

# Core interests, Crisis Behavior and Reputation: A survey experiment with national security elites

Word Count: 11,981

November 4, 2019

## Abstract

Does crisis behavior signal information about resolve or, aggressive long-term intentions, or is there too much uncertainty to signal anything at all? Studies of iterated crisis bargaining have found this question difficult to answer because they assume states draw their values for each issue independently. I argue that pre-crisis foreign policy context — whether an issue under dispute is consistent or inconsistent with the challenger’s declared core interests — determines what defenders infer. When challengers fight for a declared core interest defenders draw inferences about resolve to fight for core interest. When challengers fight over a peripheral interest, defenders draw inferences about the challenger’s aggressive strategic intentions. I test this argument with an elite survey experiment that extends a realistic war game exercises that the National Security Council participated in. The subjects — intelligence, defense and foreign policy experts in Washington DC — are randomly assign information about the challenger’s declared foreign policy and military intervention behavior, then asked to assess the challenger’s intentions and resolve. Using random assignment on an elite sample and exactly simulating an assessment task they regularly participate in, I validate my theory of foreign policy context and differential crisis signaling. I import elite survey methods from judicial, medical and business studies.

# Introduction

What inferences do states draw when they observe their rivals use force to take territory? There are two research agendas on this question that emphasize different kinds of private information. In the crisis bargaining framework, fighting signals high resolve to fight in future disputes (Sartori, 2005, 2002; Renshon, Dafoe, and Huth, 2018). In the security dilemma framework, fighting signals aggressive strategic intentions (Kydd, 2005).

These researchers all believe that a state's strategic intentions are defined relative to its resolve to fight (Kertzer, 2016). But throughout history, whether policy-makers drew inferences about a rival's resolve to fight, or strategic intentions had dramatically different implications for the patterns of war and peace that followed. For example, in 1945, Stalin orchestrated a coup in Romania. Although alarmed by the lengths Stalin would go to (Kitchen, 1986, p261), the Western Allies assessed that "Russia policy, however distasteful it may be to us... has the air of remaining a policy of limited objectives."<sup>1</sup> For a year longer, they hoped Stalin would peacefully integrate into world order. If anything, events in Romania only increased the size of concessions that the British were willing to make (Kitchen, 1986, p261). By contrast, when Stalin instigated the Iran crisis in 1946, the Western Allies concluded that Russia "will adopt a policy of opportunism to extend her influence wherever possible."<sup>2</sup> This revised assessment about Stalin's intentions led to the origins of containment.

Evidence for these different frameworks is mixed. Some report fighting in a crisis signals high resolve leading to concessions in future crises, others report it signals aggressive strategic intentions leading to preventive war, and others still find no consistent pattern (Danilovic, 2001; Huth and Russett, 1990; Huth, 1997; Schultz, 2001b; Levy, 2012; Wood, 2012; Fearon, 1994b). A reputation for one kind of private information may be activated for the duration

---

<sup>1</sup>Clark Kerr's assessment: FO371/47941 Mar. 27, 1945

<sup>2</sup>Report of the Joint Intelligence Sub-Committee of Great Britain: 'Russia's Strategic Interests and Intentions,' Mar. 1, 1946.

of leadership tenure or given certain country traits or when crises occur over “similar” issues (Renshon et al., 2018; Renshon, 2016; Tomz, 2007; Levy, McKoy, Poast, and Wallace, 2015; Uzonyi, Souva, and Golder, 2012; Weeks, 2008; Kertzer, 2017; Weisiger and Yarhi-Milo, 2015). But researchers do not understand why states sometimes draw one type of inference and not the other, if these inferences are rational (Mercer, 1996; Yarhi-Milo, 2014; Tang, 2005), or systematic (Press, 2007).

In this paper I ask two questions: why do these two frameworks produce different predictions even though they both claim to study the same behavior? When should we expect the predictions of each to bare out? I argue that these divergent predictions arise because each framework makes different assumptions about how a defender can weild force. These assumptions produce different incentives to misrepresent in otherwise similar strategic interactions. In the crisis bargaining framework, defenders decide whether to deploy force in the middle of a crisis to deter a challenger from fighting. This creates incentives for challengers to *overstate* their resolve to fight in the current crisis. Given incentives to overstate resolve, defenders infer that the challenger is highly resolved to fight tomorrow if the challenger fights today. In the security dilemma framework, defenders use their military pre-crisis to topple a challenger and prevent any future crises from materializing. This creates incentives for challengers to *understate* their resolve to fight in many future crises. When a challenger fights knowing that it may trigger major war, the defender infers that the challenger is highly resolved to fight in all future crises.

The question, then, is how do states know whether they are playing the security dilemma game, or the repeated crisis game? I argue that the strategic interaction depends on whether a crisis errupts over an issues that is consistent with the challenger’s long-standing declared core interests or not. Through history, states have identified what issues and territories were most important to them through declared peace-time foreign policies. For example, in 1823 the United States used the Monroe Doctrine to declare that Western hegemony was in its core interests. In 1841, Tsar Nicholas declared that Russia would protect Orthodox

Christians worldwide. In 1932, Hitler declared that sovereign control of German speaking territories was in Germany's core interests.

Like [Trager \(2011\)](#) I find that states can use costless core interests claims in peace-time to explain what issues they value the most. Unlike [Trager \(2011\)](#) I allow the defender to exploit historical information about a challenger to understand which tangible real-world objectives are consistent with the challenger's declared core interests ([Joseph, 2017b](#)). As a result, challengers use their declared core interests to: (1) explain that they are highly resolved to fight for specific issues (their core interests); and (2) promise that their aims are limited because they will not fight for peripheral interests.

Defenders understand that challengers have complex incentives to overstate their resolve and understate their intentions. However, defenders can combine these costless peace-time declarations with the challenger's costly decision to fight in a crisis to make nuanced inferences. When the issue in dispute is consistent with the challenger's declared core interests, then states play the crisis bargaining game. If the challenger fights, the defender infers that the challenger will fight over core interests in the future and will stay out of future crises over the challenger's declared core interests. When a crisis erupts over a challenger's declared peripheral interests, then states play the security dilemma game. If the challenger fights, the defender infers that the challenger has expansive aims and will fight in any crisis that erupts. As a result, the defender turns to preventive war.

For example, in 1903 the United States threatened British interests in Alaska. The British did not know if the United States would actually fight for concessions in Alaska if Britain refused to give in, or if the United States would make expansive future demands even if Britain made concessions. But Britain did know that American interests in Alaska were consistent with the Monroe Doctrine. When the Americans deployed forces to Alaska, the British inferred that the United States was sufficiently resolved to fight to enforce the Monroe Doctrine (a-la the repeated crisis game). But the Alaska Boundary Dispute did

not cause the British to question American ambition beyond the Western Hemisphere.<sup>3</sup> I conjecture that British inferences would have been different if the United States invaded British colonies in Africa rather than Alaska. In this case, the British would have inferred that the United States had aggressive strategic intentions (a-la the security dilemma game).

The key piece of evidence to evaluate my theory—how decision-makers process information and form beliefs following rare, highly classified national security crises—is difficult to observe. To properly identify an effect, I must measure how national security decision-makers assess a rival’s intentions and resolve pre-crisis, then randomly assign a crisis event and observe how these decision-makers updated their beliefs. I do just that through a novel experiment with national security decision-makers. My vignette replicates a “war game” exercise that the National Security Council participated in. In it, subjects—real-world intelligence and foreign policy professionals in Washington DC—assessed the intentions and resolve of an emerging threat to US national security interests. I randomly assigned information about core interest claims and military crises and observed variation in subjects’ responses to these differences.

Prior experiments demonstrate that the public opinion can change based on crisis behavior (Renshon, 2015; Kertzer and McGraw, 2012; Levendusky and Horowitz, 2012; Gottfried and Trager, 2016) and that citizens update beliefs in stylized games (Tingley and Walter, 2011; Tingley, 2011). But scholars expect a lot from decision-makers who assess complex events with uncertainty across multiple dimensions. Many question if elites are subject to specific biases that may confound their assessments (Yarhi-Milo, 2014). Advancing a handful of political experiments that analyze elites samples (Renshon, 2015; Mintz, 2004; Hafner-Burton, LeVeck, and Victor, 2016; Friedman, Lerner, and Zeckhauser, Friedman et al.), my experiment follows recent trends in behavioral economics to place expert decision-makers in settings that realistically reflect their work (see Fudenberg, 2006; Camerer, Loewenstein, and Rabin, 2004, for review). I import these design choices to shed light on national security

---

<sup>3</sup>Britain did not infer that the US was benign. Rather the Alaska dispute did not effect British assessments about American intentions.

decision-making.

My argument unites studies of uncertainty about resolve (Dafoe, Renshon, and Huth, 2014) and strategic intentions (Kydd, 2005) within the one framework. It compliments studies on the conditions under which reputation adheres by moving beyond the properties of actors and instead looking at the content of a crisis (Renshon et al., 2018; Renshon, 2016; Tomz, 2007; Levy et al., 2015; Weeks, 2008). In doing so, I can explain, for example, how the Soviets drew inferences about the United States' resolve in some cases, but also why Hopf (1994) "reports that Soviet leaders did not infer from US actions in peripheral regions anything about likely behavior in Europe or East Asia."<sup>4</sup> It also takes preferences seriously (Moravcsik, 1998) by theorizing about the link between a state's underlying foreign policy priorities and the specific values of issues and territories that may arise (Jackson and Morelli, 2011).

## 1 Competing theories for drawing inferences from crises.

Theories of crisis bargaining and uncertainty study a repeated interaction between two states: a challenger ( $C$ , she) and a defender ( $D$ , he) that vie for influence over the rest of the world. At the onset, the defender is assumed to have his preferred policy over all contested foreign policy issues. In each period, a crisis emerges over one issue and the challenger has the opportunity to fight and revise the status quo in her favor, or not fight and live with the status quo.

The defender is uncertain about how the challenger will behave because there are different kinds of issues over which a crisis can emerge (crisis-specific uncertainty), and different kinds of challengers (challenger-specific uncertainty). There is crisis-specific uncertainty because crises can emerge over different types of issues (call them issue type  $A$ , and  $B$ ), and some challengers care about some of these issues more than others. As a result, when a crisis erupts the defender does not know if the challenger values the issue in dispute high ( $\pi_H$ ) or

---

<sup>4</sup>Weisiger and Yarhi-Milo (2015)

low ( $\pi_L < \pi_H$ ) (Rosato, 2015; Sartori, 2005).

There is challenger-specific uncertainty because the defender does not know how much the challenger cares about each issue type relative to each other (Trager, 2011), or relative to the cost of fighting (cf Fearon, 1995). Challengers can hold at least three levels of resolve: A greedy (or highly resolved) type that values all issues a lot ( $C_g$ ); a benign (or weakly resolved) type that values all issues a little ( $C_b$ ); and a limited aims (or conditionally resolved) type that values some issues a lot (possible issue-type  $A$ ) and others a little ( $C_l$ ). In the simplest version, researchers summarize the challenger’s value for different kinds of issues conditional on the Challenger’s resolve as:

$$C_g : \pi_H = A = B \text{ (values both issue types A and B high)} \quad (1)$$

$$C_l : \pi_H = A > B = \pi_L \text{ (values A high and B low)} \quad (2)$$

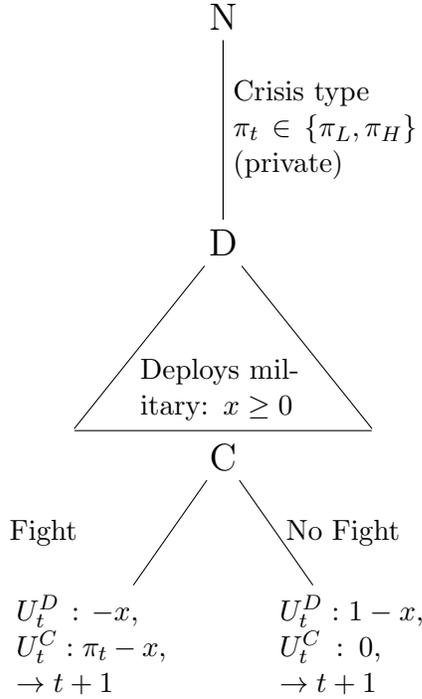
$$C_b : A = B = \pi_L \text{ (values both A and B low)}. \quad (3)$$

These equations represent that C can have different values (high/low) for different kinds of issues (A/B) depending on her underlying resolve. Different studies add to this basic framework to study how different military (Slantchev, 2003; Yoder, 2019), and diplomatic (Fearon, 1994a; Sartori, 2005) choices can help ease information problems. However, they all want to understand how C’s decision to fight in a crisis reveals information about her underlying resolve. That is, is C type:  $C_b, C_l, C_g$ ? I explicitly focus on this question: given the different pathways a crisis can take, what inference does D make about C’s resolve once D observes C fight in a crisis?

The repeated crisis framework and the security dilemma framework start with this basic question. But they each make different assumptions about how defenders use military force that alters the challenger’s incentives to misrepresent and causes defenders to focus on different information problems. To highlight these differences, Figure 1 depicts the extensive

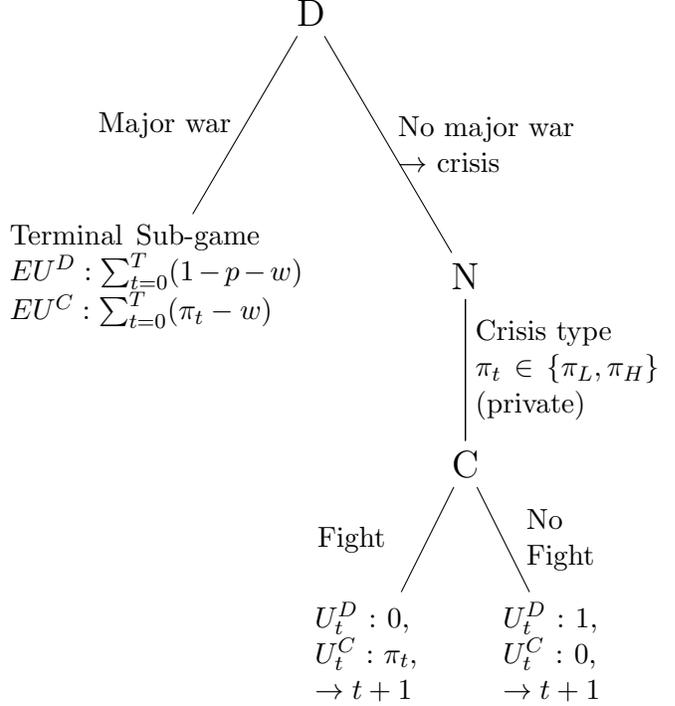
Figure 1: One period ( $t$ ) in different frameworks on crisis bargaining

Figure 2: Repeated Crisis Game



Nature chooses the crisis-type (privately for C). D chooses to commit  $x \geq 0$  military effort in case C fights. C then chooses to fight and take the issue or not. The game always moves to  $t + 1$ . D always pays the cost of militarization ( $x$ ), C only pays  $x$  if she fights.

Figure 3: Security Dilemma Game



D chooses between a game-ending major war and allowing a crisis to materialize. If D selects major war, players take war pay-offs for all future crises. If D allows a crisis to materialize, Nature chooses C's value, and C can fight and take that issue or not and the game moves to  $t + 1$ .

form of these frameworks for one period ( $t$ ), which we can think of as one crisis episode, of a repeated interaction between a challenger and a defender.

Panel (a) represents one period of the repeated crisis framework. Once a crisis emerges, the defender chooses how much to militarize in preparation for fighting if the crisis escalates (defender sets  $x \geq 0$ ). The challenger observes the defender's military deployment, then chooses whether to fight for the issue, or not fight and let D keep it.

The defender's militarization choice has short-term effects: it raises the challenger's cost of fighting in the current crisis but does not effect choices in future crises. The defender wants to militarize to deter the challenger from fighting. However, challengers that value

the issue more than the cost of fighting will fight no matter what the defender does. Against these challengers, the defender prefers to concede the issue outright, rather than concede the issue and pay the militarization cost (Sartori, 2005).

These short-term effects create incentives for challengers to *overstate* their resolve to fight in the current crisis. If C can convince D she cares intensely about the current crisis, D will not militarize because he believes C will fight no matter what he does. Defenders exploit large military deployments to screen out challengers with low resolve (Slantchev, 2003).<sup>5</sup> When the challenger fights in one crisis, it causes the defender to increase his confidence that the challenger is resolved to fight in future crises. However, if the challenger backs down, the defender decreases his confidence that the challenger will fight in future disputes (Renshon et al., 2018).

**Existing Expectation 1** *When a challenger faces incentives to overstate her interest in the next crisis but fights anyway, the defender increases his confidence that the challenger is resolved to fight for issues the challenger values high.*

Panel (b) represents the basic set-up of the security dilemma framework. In this framework, defenders use their military to fight major wars designed to overthrow the challenger and prevent future crises from materializing.<sup>6</sup> Since major war has implications for future periods, the defender is not only worried about what the challenger will do in the current crisis but rather how the challenger will behave in future crises (Glaser, 2010). When challengers are likely to fight for many future issues, the defender prefers to fight one large war in the present to prevent a barrage of iterated threats and concessions (Streich and Levy, 2007; Powell, 1999). When challengers worry about the threat of major war, they *understate* their interest in fighting for many future crises to avoid it (Jervis, 1978). However, highly resolved (greedy) challengers are tempted by the short-term gains and fight anyway.

---

<sup>5</sup>In advanced models D combines large military deployments with diplomacy (Sartori, 2002), audience costs (Fearon, 1995) or other mechanisms (Schultz, 2001a) to screen out low-resolved challengers.

<sup>6</sup>This is usually modeled as a costly lottery that C wins with probability  $p$ . The winner of war automatically wins all future crises and both players pay a cost  $w$ .

**Existing Expectation 2** *When a challenger faces incentives to understate her strategic intentions but fights in a crisis anyway, the defender increases his confidence that the challenger will fight opportunistically in many future crises.*

In the real-world both of these dynamics work simultaneously. Defenders can choose (1) major war to prevent a crisis from materializing, and, once a crisis looms, (2) to deploy forces to raise the challenger's cost of taking the current issue in dispute.

A third group of researchers argue that multiple options for competition, multi-dimensional uncertainty and complex incentives to misrepresent make signaling resolve too difficult. Defenders do not know if the challenger's decision to fight signals expansive aims, or that the challenger has limited aims and cares a lot about the issue in dispute (Press, 2007). Since crises are high-stakes and uncertain interactions (Rosato, 2015), and that Defenders care so much about the issues they are contesting, Defenders will not risk losing on an issue to gather information about the challenger's type (Mearsheimer, 2001; Press, 2007).

**Existing Expectation 3** *In the high stakes and uncertain world that Press describes, if the challenger has complex incentives to mis-represent but fights anyway, the defender should expect the worst.*

## **2 The mediating variable: Declared core interests.**

I argue that defenders rely on the challenger's claims about her core interests to systematically identify high and low-valued issues before a crisis emerges. When a crisis erupts, the inferences that defenders draw from a challenger's crisis behavior depend on whether a crisis erupts over an issue that is consistent with the challenger's declared core interests or not. When a crisis erupts over a challenger's declared core interest, states play the repeated crisis game. If the challenger fights, the defender infers that the challenger is resolved to fight for core interests. However, the defender does not assume that the challenger is more likely to be greedy. When a crisis erupts over a declared peripheral interest, states play the

security dilemma game. If the challenger fights, defenders infer that the challenger is greedy, and willing to fight for all future issues.

First, I define a challenger's declared core interests. Second, I explain the strategic implications from integrating core interest claims into crisis bargaining.

## 2.1 Declared Core Interests

I define a challenger's core interests as the tangible objectives that she will use military force to fight for even in the face of resistance from the defender. I refer to issues that the challenger is unwilling to fight over as its peripheral interests. I am interested in the challenger's declared core interests rather than her actual interests. That is, the issues and territories that the challenger claims she is willing to use military force to take, control and maintain based on foreign policy speeches and documents she produces in peacetime (i.e. before the first crisis).

Many states declare their core interests through publicly available national security documents. For example, Australia's Defense White Paper identifies Australia's core interests as political freedom and stability for nation-states in the Pacific Ocean, honoring its alliance commitments to the United States and protecting Australian trade interests. In other states, leaders use public speeches or diplomatic representations to declare their core interests.

To be clear, it is costless for challengers to declare their core interests, and they face incentives to misrepresent that are especially strong once a crisis emerges.<sup>7</sup> But challengers cannot easily alter their core interest claims in the midst of a crisis. The reason is that a state's declared core interests are not a list of random issues they claim to care about. Rather, a state's core interests are tied together by an underling principle that motivates her foreign policy (Joseph, 2017b). Some claim that they are motivated by restoring their historical borders (Carter and Goemans, 2011), others claim international security (Waltz,

---

<sup>7</sup>My results would only be strengthened if states paid costs (such as audience costs) to declare their core interests. I chose costless messages because most analysts believe that these White Papers (which often stay consistent across different administrations) are largely costless for leaders to develop.

1979), status (Gilpin, 1983; Renshon, 2016), ethnic-nationalism or human rights/democracy promotion.

When challengers make their foreign policy objectives public, they justify them in terms of these principles to explain which issues are core interests and which are not (Joseph, 2017b). For example, in 1830, Tzar Nicholas declared that Russia's core interest was to defend the autonomy of territories populated by orthodox Christians. He did not list out all of the conditions under which he would fight, or the territories he would protect. Other European powers did not know if Nicholas' core interest claim was genuine. What they did know, generally speaking, was where all the Orthodox Christians lived. When the Crimean Crisis erupted in 1836, Nicholas had not specified that he would fight in the Crimea. Yet, the British knew that this crisis was consistent with his claim to protect Orthodox Christians because many Christians lived in Crimea. Therefore, it was easy for Nicholas to claim he would fight for this issue and still promise he would not fight for all issues.

## 2.2 Strategic Implications

I argue that core interest claims mediate the conditions under which crisis bargaining produces the predictions of the repeated crises or security dilemma frameworks. However, the predictions I draw ultimately follow from these well studied, and extensively formalized strategic interactions. I summarized these results in the Existing Expectations above. Rather than re-hash these well-specified mechanisms in the manuscript, I describe the mediating role that core interests play informally and derive a formal argument in Appendix A.

I accept there may be psychological or social mechanisms that explain why core interests claims play this mediating role.<sup>8</sup> I develop a rational argument because I want to integrate my theory into the rationalist crisis bargaining literature. In particular, I want to show that my predictions adhere even in the high stakes and uncertain conditions that Press (2007) describes. To do that, I solve for a model where defenders have information about a

---

<sup>8</sup>If true, the strategic discussion below is unnecessary and I can move straight to hypotheses.

challenger's core interest. I focus explicitly on parameter ranges that match the high-stakes and uncertain conditions that Press describes.<sup>9</sup>

Formalization also allows me to draw precise predictions about the defender's beliefs following the resolution of the first crisis. I use the formal model to compute D's beliefs about C's resolve ( $pr(C = C_g, C_l, C_b)$ ) at the end of the first crisis. My hypotheses follow from comparing D's posterior beliefs about C's resolve at the end of the first crisis given: (1) C's core interest claim; (2) whether the crisis is over a core or peripheral interest; (3) whether C fought in the first crisis.

Finally, I can exploit the formal model to contrast my theory directly with the existing literature. To do this, I report D's beliefs from a counter-factual model where defenders can use force in multiple ways, but challengers cannot declare their core interests. In this model, I show that (Press, 2007)'s predictions bare out. I develop two additional hypotheses that compare D's beliefs when C fights for a declared core interest (my model) and when C fights but core interests are undefined (the conventional wisdom).

Figure 4 visualizes how I integrate core interests into a crisis bargaining model. As in the baseline model, there is crisis-specific uncertainty and challenger-specific uncertainty. The defender does not know the challenger's underlying level of resolve  $C_j \in \{C_b, C_l, C_g\}$ ; or which issues the challenger values high or low.

Unlike the baseline model, I replace the generic issue types over which a crisis can erupt (A, B) with specific issue types: ethnic and security issues ( $e, s$ ). As a result, the challenger can hold core interests in either ethnic or security issues  $C_i \in \{C_e, C_s\}$ . The inclusion of core interests creates two new sources of information. First, the defender observes whether a crisis erupts over an issues that relates to the challenger's ethnic or security interests ( $e, s$ ).<sup>10</sup> Second, before the first crisis begins, challengers declare their core interests through a costless message.<sup>11</sup>

---

<sup>9</sup>Focusing on other parameter ranges only improves my results.

<sup>10</sup>To be clear, the defender does not know which issue type is the challenger's core interest. But the defender does know which issue type a crisis is over.

<sup>11</sup>Substantively, this is like writing a Defense White Paper. This message is unlike Sartori (2005) because

Figure 4: Core interests and the crisis bargaining game.

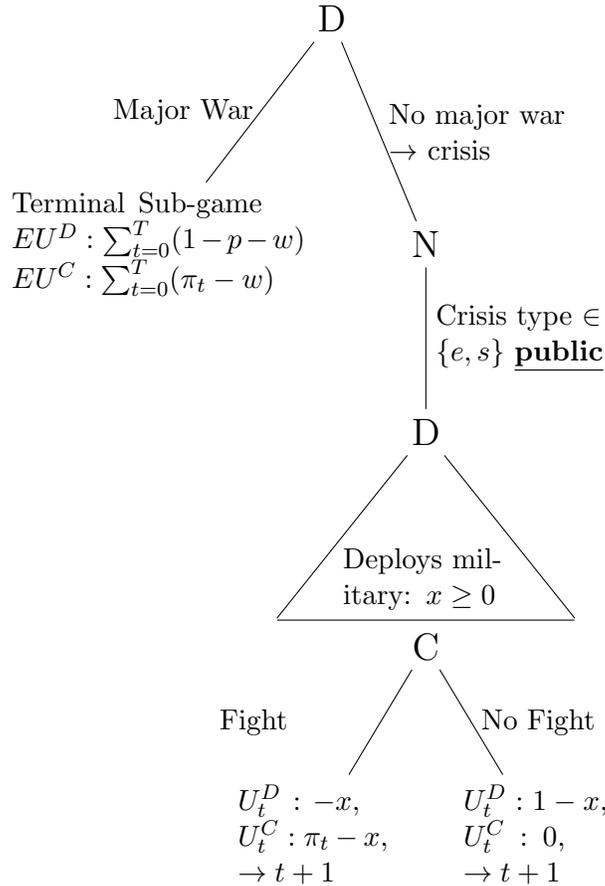
1. A new, pre-crisis stage:

a. Nature assigns:

- C's resolve:  $C_i \in \{C_b, C_l, C_g\}$
- C's core interests:  $\pi_H \in \{e, s\}$

b. C observes type (private), sends costless message about her type to D.

2. A re-worked crisis period when the nature of crises are observable:



Combining past frameworks, I allow D to choose either major war pre-crisis, or militarization in the middle of a crisis. Unlike past frameworks, I assume: (1) nature chooses the type of crisis— $e, s$ —that emerges and not just whether C values it high or low; and (2) Nature reveals the type of crisis publically to both players and not privately to C.

These new information sources create four different pathways the game can take leading up to the resolution of the first crisis. A crisis can erupt over C's declared core interests and C can choose to fight or not. Alternatively, a crisis can erupt over C's declared peripheral interests and C can choose to fight or not.<sup>12</sup>

Table 1 summarizes D's beliefs about C's resolve at the end of the first crisis given the four possible pathways the game can take. Unlike the conventional wisdom, D does not always expect the worst if C chooses to fight. If C chooses to fight over a declared core interest, then D ends the first crisis more optimistic that C has limited aims (is type  $C_l$ ), and will only fight for core interests in the future, than D was at the very beginning of the game. However, if C fights for a declared peripheral interest, D becomes certain that C is the greedy type, who will fight in all future disputes.

Two factors drive these differences. First, similar to the mechanisms studied by [Trager \(2011\)](#) and [\(Joseph, 2017b\)](#), a challenger's core interest claim is partially informative even though it is cheap-talk. This is somewhat surprising because there are complex incentives for challengers to misrepresent: greedy types ( $C_g$ ) want to understate their total aims, and benign types ( $C_b$ ) want to overstate their resolve to fight in a specific crisis. This produces an incentive for all types to promise that they hold limited aims ( $C_l$ ). The reason is that limited aims types can credibly promise to fight for a few specific issues (and so D should stay-out in a crisis over those issues), but they can also credibly promise to only fight for those specific issues (and so D should not enact major war).

But challengers cannot just say they have limited aims. They also have to specify what those limited aims are by identifying specific core interests. Since greedy and benign types value both security and ethnic issues about the same, and they do not know what crisis is going to erupt, they are indifferent about which core interests they claim to care about. They randomly choose which core interest to pretend they value the most and hope that

---

it occurs before the challenger knows what issue the first crisis will be over. If, like Sartori, I assumed the challenger observed a crisis, then made a core interest claim I would not derive the same predictions.

<sup>12</sup>In the standard model, there are only two pathways the game can take: the challenger chooses to fight or not.

Table 1: D’s beliefs about C’s resolve at the end of the first crisis

	Benign ( $C_b$ )	Limited ( $C_l$ )	Greedy ( $C_g$ )
Prior	1/3	1/3	1/3
Fight for core	0	2/3	1/3
Fight for peripheral	0	0	1
Does not fight for core	1	0	0
Does not fight for peripheral	1/3	2/3	0

Rows are C’s first crisis choice. Columns are D’s beliefs about C’s type given C’s first crisis choice. Beliefs when C chooses not to fight are reported in Appendix A.

opportunities for crises emerge that will lead to uncontested concessions.

In contrast, challengers that actually hold limited aims do value one issue-type more than the other. The limited aims challenger wants to fight for her core interests at all costs. However, she is also willing to forgo her peripheral interests to avoid preventive war. Thus, she faces additional incentives to honestly reveal her core interests.

Second, once a crisis emerges, D can condition his military deployments on whether the issues matches C’s declared core interests. If a crisis erupts over a declared core interest, D sets  $x$  high to rule out benign types.<sup>13</sup> If a crisis erupts over a peripheral interest, D sets  $x$  low. Instead, D relies on a credible threat of preventive war if C fights to deter limited aims types from fighting.<sup>14</sup>

This information has cumulative effects that influence D’s strategy in the second period. At the end of the first crisis, D knows both the nature of the first crisis, and also whether C fought. But once the second crisis starts, D also knows if the second crisis is over an issue that was similar or different to the first crisis. If the second crisis erupts over the same kind of issue, D can effectively exploit what he learnt in the first crisis because D knows how C behaves in that sort of situation. For example, if C fights for a core interest in the first crisis. D knows that C will fight for a core interest in the second crisis as well.

These mechanisms allow D to make different inferences based on C’s core interest claims, whether the first crisis was over a declared core interest and whether C fought or not. In

<sup>13</sup>Following the strategy identified by (Sartori, 2005; Renshon et al., 2018) and others.

<sup>14</sup>Following the strategy identified by (Kydd, 2005) and others.

what follows I focus on the case where C chooses to fight. To test the mediating role that core interest claims play, I draw hypotheses about the different inferences that D makes depending on whether or not C fought for an issue that was consistent or inconsistent with her declared core interests:

**Hypothesis 1** *When a challenger fights for an issue that is consistent with her declared core interest, defenders are more confident that the challenger is the type that will only fight for her core interests ( $C_1$ ) than they would have been if the challenger fought for her peripheral interests.*

**Hypothesis 2** *When a challenger fights for a declared peripheral interests, defenders are more confident that the challenger is the greedy type ( $C_g$ ) that will fight opportunistically whenever the opportunity presents itself than they would have been if they observed the challenger fight for her declared core interests.*

These hypotheses are derived from a world in which

These hypotheses follow in a world where D can identify which crisis issues are consistent with C's core interest claims. By contrast, the existing literature assumes that challengers cannot declare their core interests, and defenders have no information about the nature of issues in a crisis (Sartori, 2005; Kydd, 2005; Rosato, 2015, cf).<sup>15</sup> I formalize this conventional wisdom in Appendix B and focus on the same high-stakes and uncertain parameter ranges I used to produce D's posterior beliefs in Table 1. In this counter-factual model, I find a unique equilibrium consistent with the predictions of Press (2007) (existing expectation 3). In equilibrium, if C fights in the first crisis, D updates his beliefs such that D is now much more confident that C is greedy (highly resolved) type. Based on this heightened level of confidence that C is greedy, D chooses major war to prevent the next crisis from materializing.<sup>16</sup>

---

<sup>15</sup>Even in Sartori (2005); Fearon (1994a) challengers do not declare their core interests in peace-time. Rather, they declare their interest in an issue once a crisis has started.

<sup>16</sup>For a clear informal explanation of the mechanism, see Press (2007).

Table 2: D’s beliefs about C’s resolve in a world where core interests are undefined

	Benign ( $C_b$ )	Limited ( $C_l$ )	Greedy ( $C_g$ )
Prior	1/3	1/3	1/3
C fights in first crisis	0	1/3	2/3
C does not fight in first crisis	2/3	1/3	0

Each rows reports D’s beliefs about C’s resolve at the end of the first crisis given the different pathways that the game could take. Since D has no information about the nature of the crisis, then D only observes if C fights or not. Posterior beliefs from formal analysis in Appendix B.

In Table 2, I report D’s beliefs about C’s resolve at the end of the first crisis depending on whether or not C fought in the counter-factual model. I can contrast the results from my framework (Table 1) where D has information about C’s core interests and C fights, with the results from the existing literature, where core interests are undefined and C fights (Table 2).

**Hypothesis 3** *When a challenger fights for an issue that is consistent with her declared core interest, defenders are more confident that the challenger is the type that will only fight for her core interests ( $C_l$ ) than they would have been if the challenger fought but did not have clearly defined core interests.*

**Hypothesis 4** *When a challenger has no declared core interests and chooses to fight, defenders are more confident that the challenger is the greedy type ( $C_G$ ) that will fight opportunistically whenever the opportunity presents itself than they would have been if they observed the challenger fight for her declared core interests.*

## 2.3 Re-visiting existing evidence

Past studies have reported evidence about how defenders draw inferences from, and respond to, a challenger’s decision to fight in a crisis. In this section I re-interpret this evidence in the context of my theory. I analyze archival studies about Anglo-German relations in the 1930s and quantitative studies. I do this to demonstrate that I can clarify seemingly contradictory findings about crisis and reputation as conditional effects. Exploring past studies

also demonstrates that my predictions generalize to observed patterns across different types of data.

Archival researchers have asked why British elites chose to appease Hitler despite Hitler's repeated use of force to take territory. Although the details vary, scholars identify crises—for example, the re-militarization of the Rhineland and the Anschluss—as key events that should have led the British to conclude that Hitler had vast aims. Since the British did not update their beliefs, these studies conclude that either British elites were irrational, naive, or that defenders should not update based on crisis behavior (cf [Yarhi-Milo, 2014](#); [Gilbert, 1972](#); [Barnett, 1986](#)).<sup>17</sup>

Yet British assessments are well explained by mapping these crises onto Hitler's declared core interests. The British believed that Hitler may have been motivated by uniting ethnic-Germans under one government. Each of these crises is consistent with that basic principle. For example, when Hitler re-militarized the Rhineland, the British War Secretary told the German Ambassador, “through the British people were prepared to fight for France in the event of a German incursion into French territory, they would not resort to arms on account of the recent occupation of the Rhineland. The people did not know much about the demilitarization provisions and most of them probably took the view that they did not care ‘two hoots’ about the Germans reoccupying their own territory.”<sup>18</sup>

The shift in British assessments following the failure of the Munich Agreement are also consistent with my theory. In 1934, Hitler mobilized forces to take Czechoslovakia. Alarmed by these events, the British attempted to mediate a settlement. Ultimately, Hitler accepted territorial concessions over only the German speaking parts of Czechoslovakia. Based on this compromise, the British were optimistic about peace in Europe. It was only after Hitler took the non-German speaking parts of Czechoslovakia that the British prime minister became

---

<sup>17</sup>One exception is [Ripsman and Levy \(2008\)](#). Their account is consistent with mine. As they point out, the most senior British elites were, at best, uncertain about Hitler's strategic intentions up until Hitler violated Munich.

<sup>18</sup>Quoted in [Weinberg \(1980\)](#) p259.

convinced that Hitler's aims were expansive.<sup>19</sup> The failure of the Munich Agreement was not the first time that Hitler had taken territory, threatened force, or even violated an international agreement. It was, however, the first time that Hitler took territory that was inconsistent with his declared core interests.

Researchers that furnish quantitative evidence also reach divergent findings. [McManus \(2014\)](#) finds that American presidents who make forceful threats are more likely to receive favorable conflict outcomes because they can generate a reputation for high resolve.<sup>20</sup> In contrast, [Gottfried and Trager \(2016\)](#) find that “aggressive foreign rhetoric makes it harder for a leader to offer a dramatic settlement involving substantial concessions to preserve the peace,” and also “dramatically increases the approval of presidents who prosecute successful wars.” Different still, [Wood \(2012\)](#) finds no relationship between threats and conflict outcomes across crises.

At first sight, the evidence seems contradictory. My theory clarifies why these scholars observe different effects. McManus codes the severity of all American presidential demands since 1950 using a machine-coded dictionary. The dictionary relies on pre-coded phrases as high or low resolved. As a result of this procedure, she codes both statements to retaliate against those that kill American soldiers, and the Cuban Missile Crisis as high-resolved threats. In contrast, she codes both statements about the 1991 Gulf War and the Soviet withdrawal of Afghanistan as low-resolved demands. At least these archetypal examples demonstrate a correlation between strong threats and core interests. Consistent with my theory, she finds that Americans can rely on threats to demonstrate resolve to fight when they make threats over their core interests.

[Gottfried and Trager \(2016\)](#) report a survey experiment based on an ongoing Russo-American dispute over oil-rich arctic territories. Participants are told that Russian interests in the arctic are not related to Russia's core interests. Therefore, subjects that observe

---

<sup>19</sup>This account is consistent with ([Ripsman and Levy, 2008](#); [Bullock, 1971](#); [Hillgruber, 1974](#))

<sup>20</sup>This debate centers on threats, but as [McManus \(2014, p728\)](#) notes, “the logic should apply to both conflicts that are decided without the use of force and conflicts in which force is used.”

Russia threaten force to take this territory, observe Russia threaten a peripheral interest. Consistent with my theory, when Russia fights for a peripheral interest subjects are more likely to support the use of force against Russia at the first available chance and are less likely to support short-term concessions.

These findings demonstrate that my theory is consistent with a wide variety of evidence. When a conflict arose over a challenger’s core interest, archival research showed that the defender inferred that the challenger was resolved to fight for core interests. The statistics showed that defenders were willing to make more concessions in future disputes. In contrast, when a conflict arose over a peripheral interests, archival research found that the defender inferred that the challenger was greedy. The statistics showed that challengers were willing to fight at the first available opportunity.

### 3 Causal Evidence

I present evidence from an elite survey experiment on 93 foreign policy, intelligence and defense professionals. The survey presented information about a fictional challenger named Bandaria. Subjects were randomly assigned information about Bandaria and then asked to assess Bandaria’s intentions and resolve to fight over core or peripheral issues. The hypotheses and research design were pre-registered at [removed-for-review].

I chose a survey experiment rather than an observational study for two reasons. First, my theory partially drew from seemingly divergent results from existing empirical research. Even though my mechanism well explains these divergent findings, I partially induced it from prior research. Thus, relying on this prior evidence to test my theory exposes me to validity concerns.

Second, I must overcome endogeneity concerns that follow from the challenger’s incentives to strategically manipulate her core interest claims. To identify an effect of declared core interest, I need to measure each challenger’s actual core interests, and defender’s priors about

them, then apply a selection model to explain the core interest claims that challengers make. Given that challenger’s national security interests and defender’s beliefs are highly classified, difficult to observe and hotly contested in the literature, I do not think it is possible to measure these variables with high confidence.

Through generating new data that randomly assigns declared core interests and military interventions I can causally identify an effect. This builds a bridge between my theoretical framework and the divergent findings from observational data that my theory well explains.

### **3.1 A Scenario-Based, Elite Survey Experiment**

This survey departs from most other survey experiments in international relations in two ways. First, subjects were foreign policy elites. Second, the vignette was a detailed hypothetical scenario that closely reflected a real-life war game exercise that national security professionals participated in, rather than a stylized game or a short vignette. Below I explain the advantages of this approach, and the design choices that help me overcome its shortcomings.

To understand how elites evaluate crisis behavior I focused my attention on a convenience sample of elites. Subjects were eligible if they had briefed a Deputy Assistant Secretary, Congressperson or similarly ranked official on foreign policy issues. Subjects were asked sample inclusion questions at the end of the survey to ensure they met the elite sample frame.<sup>21</sup> 139 subjects answered at least one question, 131 completed the survey, and 93 passed attention checks. I analyze these 93 responses below. My sample is a good proxy for high-level elites for two reasons. First, all participants are successful, political officials focused in foreign affairs. It is precisely this group of people that cabinet members are drawn from. Second, subjects were selected because they provide information to senior decision-makers. High-level elites rely on facts and analysis that they receive from people in this sample. Thus, the sample has considerable influence in shaping the information that their

---

<sup>21</sup>See Appendix D.1.2 for solicitation information and D.5 for balance tests.

superiors see.

In international relations research, scholars claim that personal experiences distinguish foreign policy elites' decision-making from the general population's (cf [Saunders, 2011](#)). If foreign policy elites think differently, it is difficult to draw strong inferences about their behavior from experiments administered to a general population. In other contexts, it has been shown that professionals with specialized expertise approach their work differently than an average educated adult would the same task. For this reason, behavioral researchers increasingly turn to convenience samples to identify effects for medical doctors ([Arber, McKinlay, Adams, Marceau, Link, and O'Donnell, 2006](#); [Feldman, McKinlay, Potter, Freund, Burns, Moskowitz, and Kasten, 1997](#)), CEOs ([Rashad Abdel-Khalik, 2014](#); [Cen and Doukas, 2017](#); [Lieb and Schwarz, 2001](#)), or lawyers and judges ([Redding, Floyd, and Hawk, 2001](#)), rather than a representative sample of educated adults. These convenience samples often derive consistent result in repeated experiments on elite samples ([Redding et al., 2001](#)).

Drawing from this research, I took four steps to increase my confidence that the results are not an artifact of my sampling method. First, I solicited elites using two distinct sampling techniques that I describe in [Appendix D.1](#). Each sampling method had its own link to an identical survey. I demonstrate that the treatment effects hold controlling for the different sampling methods in [Table 4](#). Thus, I can say with confidence that one method of sampling did not determine the results because I get the same results using different sampling techniques on different sub-populations of elites. Second, I collected biographical information on president Trump's current NSC and president Obama's final NSC. [Figure 10](#) reports the summary statistics for my sample (panel a) broken out by sampling procedures and the real NSC staff (panel b) broken out by president. The variation across my sampling frames is consistent with variation in real-world NSC selection. Third, I report metadata on response attributes and attrition rates recommended by [Eysenbach \(2004\)](#) in [Appendix D.2](#). Fourth, I conducted pilot surveys on Mechanical Turk to test features of the vignette recommended by [Steiner, Atzmüller, and Su \(2017\)](#). The results are supportive.

The scenario based vignette also draws from recent innovations in behavioral economics, medicine and law. Increasingly, researchers that survey elites use scenario based exercises rather than stylized games with precise numerical payoffs (cf Collett and Childs, 2011; Arber, McKinlay, Adams, Marceau, Link, and O'Donnell, 2004). The reason is that elites make judgments in complex strategic environments that cannot be captured in stylistic games. In the national security setting, challengers are not simply profit maximizers that respond to well defined purchasing choices (like individuals in markets may). There are several dimensions of preferences and outside options that may effect decision-making.<sup>22</sup> As a result, stylistic choices do not well reflect the complex assessment process that leaders face when they assess their rival's military behavior.

National security experts are better suited to hypothetical scenarios than the average American or even other groups of experts because they participate in hypothetical war game exercises as part of their daily work. Real war plans<sup>23</sup> and National Security Estimates<sup>24</sup> are informed by war games. The vignette I developed took features from real war games that national security experts had participated in. After I developed the vignette, I received review from five foreign policy experts including a former Deputy Director of an intelligence agency to make sure the information in the vignette was consistent with the scenarios that foreign policy experts use.

One concern with these hypothetical vignettes is that policy-makers choose policies not only based on their beliefs about their rivals, but on complex inter-agency dynamics and select incentives (Allison and Zelikow, 1971). These do not apply in my case because my dependent variable is beliefs, not policy choices. Beliefs form independent of interests. In the instrument, I am careful to ask subjects about their beliefs in this scenario, rather than actions they would take or policies that they would recommend. A second concern is that subjects do not take hypothetical scenarios seriously. In Appendix D.4 I describe design

---

<sup>22</sup>Recall that main fissure in the crisis bargaining literature arises because defenders can use force in different ways, and this produces different strategic interactions.

<sup>23</sup>That is, the US military's specific plans to invade other countries

<sup>24</sup>That is, the intelligence community's assessments of foreign threats.

features, attention checks, and meta-data that demonstrate the subjects took the vignette seriously.

I chose a hypothetical scenario over a real example (e.g. a China scenario) because I did not want subjects to import outside information about a case into their answers. If subjects thought the scenario was about China, for example, they may have started the experiment with specific beliefs about the history of China’s crisis behavior. I describe design features that ensured subjects did not systematically invoke a historical case and responses that demonstrate those measures were effective in Appendix D.6.

### 3.2 Vignette

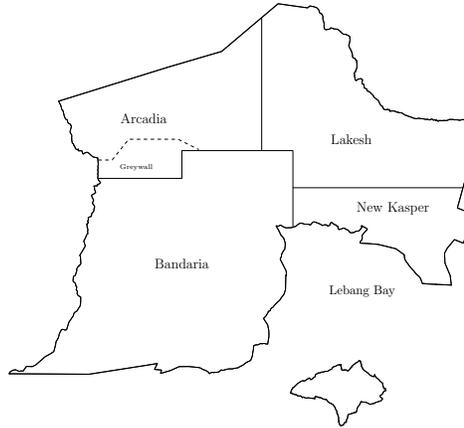
The survey instrument, presented in Appendix C, has three phases. In each phase, subjects are presented with new information about Bandaria then asked standardized questions about Bandaria’s intentions and resolve. Subjects also write text responses.

*Phase 1* provides all subjects with the same prompt and baseline information. Subjects are told that Bandaria is an emerging world power, and the American president will soon meet the Bandarian prime minister. Subjects are asked to provide the president an assessment of Bandaria’s long-run intentions and resolve in preparation for that meeting. The baseline includes information about the trajectory of Bandaria’s military spending and economic growth, socio-political environment, trade and IGO networks and geopolitics, and a map of Bandaria and its surrounding countries depicted in Figure 5. Subjects are told that there are no significant natural resource deposits.

The baseline provides two plausible dispute areas in a discussion of Bandarian regional interests. First, Bandaria’s international security is vulnerable to port closures in the Lebang Bay. These ports are controlled by New Kasper. Second, poorly treated ethnic Bandarians live in a neighboring country (Arcadia). Crucially, there is no information about which of these issues Bandaria cares about the most.

The detailed baseline addresses many confounding concerns raised by [Dafoe, Zhang, and](#)

Figure 5: Bandaria and its Neighbors



Caughey (2016). Further, interviews with senior intelligence officials in preparation for the survey suggests that the level of detail was necessary to make the scenario realistic.

In *phase two*, subjects are randomly assigned into a *core interest treatment* where Bandaria explains it is motivated by either: (1) security interests, or (2) ethnic interests. I also include (3) a counter-factual condition, that reflects the case where Bandaria does not provide any information about its core interests. In the counter-factual condition, the Bandarian prime minister talks about confidence building measures and Bandaria’s general interest in peace.

I include two core interest treatments to make sure the revelation of core interests generally and not the particular issue (security or ethnic based grievances), is doing the causal work. As an example, the ethnic treatment is:

In a private meeting, the American president asked the prime minister of Bandaria to explain Bandaria’s military spending. The Prime Minister replied: “Grave injustices have been done to ethnic Bandarians. We have a long history of supporting our Bandarian brothers in Arcadia. Ethnic-national concerns motivate our military pol-

icy.” He then said, “Of course we want to resolve this issue peacefully. But Arcadia does not realize just how concerned we are about our ethnic kin. We will use any means necessary to ensure our ethnic kin are well governed.” He continued, “Once our ethno-nationalist goals are assured, we have no reason to expand our military. All of our other foreign policy and regional concerns are less important and can be managed through UN participation, diplomacy and negotiation.”

Experts note that ethnic nationalism concerns have been central to Bandarian foreign policy over the past 10 years. Bandarian elites referred to ethnic-nationalism in private diplomatic conversations and public speeches consistently over the past 10 years.

The language modeled on de-classified minutes, letters and cables that described conversations between British elites, German Kaiser Wilhelm (1866), US president McKinley (1898), German Chancellor Hitler (1934) and Soviet premier Stalin (1932). Although the language may strike the reader as direct, it is common through history.

In *phase three*, subjects are randomly assigned into a *military intervention* treatment where Bandaria annexes: (1) territory that surrounds the Lebang Bay in New Kasper, or (2) Greywall. One corresponds with ethnic interests the other with security interests.

This treatment takes the form of breaking news. For example, the ethnic treatment is:

**Breaking News:** The Bandarian military occupied Greywall in Arcadia. Greywall is populated by ethnic Bandarians. The move comes after months of political unrest in Arcadia. The Bandarian prime minister announced plans to annex Greywall but promised fair treatment and reparations for aggrieved Arcadian citizens and businesses. The Bandarian Prime Minister insists that these events are entirely consistent with Bandaria’s interests long known to the rest of the world and Bandaria remains committed to peace and stability generally.

The treatment groups that follow are depicted in Table 3. Letters represent (s)ecurity and (e)thnic treatments, and a (c)ounter-factual. Lowercase letters are the diplomatic treatments in phase 2. Uppercase letters are the military intervention treatments in phase 3. By the end of the experiment there are 6 distinct treatment groups. However, I am mainly interested in the difference between consistent and inconsistent core interests and military intervention. To that end, I pool consistent ( $eE+sS$ ) and inconsistent ( $eS+sE$ ) groups in the main analysis.<sup>25</sup>

---

<sup>25</sup>Table 10 demonstrates that the effects hold when I consider the nationalist message and security message separately

By using two different types of declared core interests, I make sure that subjects' prior beliefs about ethnic or security issues are not driving the result.

Table 3: Treatment Groups

	Ethnic Military Dispute	Security Military Dispute
Ethnic Core Interest	$eE$ : Consistent Ethnic	$eS$ : Inconsistent
Security Core Interest	$sE$ : Inconsistent	$sS$ : Consistent Security
Counter-factual: Core undefined	$cS$ : Counter-factual	$cE$ : Counter-factual

### 3.3 Analysis plan

Subjects were asked standardized questions after each phase that reflect the hypotheses of my theory. Subjects were prompted by: “What is the percentage probability that the following statements are true?” Responses were recorded using a slide rule from 0% to 100% that moved in 5% increments. I chose this response method rather than a 7-point index for reasons that are peculiar to the subject pool. Each Agency uses a different lexicon to describe probabilities. The CIA uses a confidence scale, Hill staffers and diplomats describe probabilities with no official standard. I did not want to favor one group over another. Second, there is much debate about what confidence levels mean. For some, the level of confidence refers to the primary source material. Thus, low confidence that an assertion is true, can refer to either the credibility of the source, or that the assertion is false. I chose 5% increments because some critical numbers move along that scale. But I wanted to avoid trivially small choices that would distract subjects.<sup>26</sup>

Since I randomized treatment on a non-random sample, I report the p-value derived from the permutation test of group means suggested by [Strasser and Weber \(1999\)](#). The test identifies how confident a researcher can be that the treatment had a causal effect

<sup>26</sup>See [Lupton and Jacoby \(2016\)](#) for best practices on feeling thermometers.

on the responses of a non-random sample conditional on the responses observed and the independence of the in-sample randomization. Among the many advantages of permutation tests is that inferences are robust to small sample sizes.<sup>27</sup>

Hypotheses 2-5 analyzed the mediating effects of core interest claims assuming that hypothesis 1 was true. I test these hypotheses using two questions:

**A:** Bandaria will use military force to achieve its core foreign policy objectives.

**B:** Bandaria will use military force to expand its borders whenever the opportunity presents itself.

Hypothesis 1 and 3 focus on beliefs about whether the challenger will fight only for her declared core interests. I therefore focus on responses to question **A**. To test hypothesis 1, I compare subjects that observed Bandaria fight for core interests ( $eE, sS$ ) and peripheral interests ( $eS, sE$ ). I expect that:

$$\mu(Consistent) > \mu(Inconsistent) \tag{4}$$

To test hypothesis 3, I compare those who observed a consistent treatment ( $eE, sS$ ) to those that were not told Bandaria had well-defined core interests ( $cS, cE$ ). I measure the mean differences between how these subjects responded to **A**. I expect that:

$$\mu(Consistent) > \mu(Counter - factual) \tag{5}$$

Hypothesis 2 and 4 focus on whether or not Bandaria will fight opportunistically. I therefore focus on responses to question **B**. To test hypothesis 2, I compare those who received an inconsistent ( $eS, sE$ ) to those that received a consistent treatment ( $eE, sS$ ). I expect that:

---

<sup>27</sup>For an explanation of why they are superior to t-tests see [Ludbrook and Dudley \(1998\)](#). The p-values are interpreted like those from t-tests. Using a t-test instead of the permutation test improves the result reported in the paper.

$$\mu(\textit{Inconsistent}) > \mu(\textit{Consistent}) \tag{6}$$

To test hypothesis 4, I compare subjects that observed a consistent treatment ( $eS, sE$ ) to those that were not told Bandaria had well-defined core interests ( $cS, cE$ ). I expect that:

$$\mu(\textit{Counter - factual}) > \mu(\textit{Consistent}) \tag{7}$$

## 4 Results

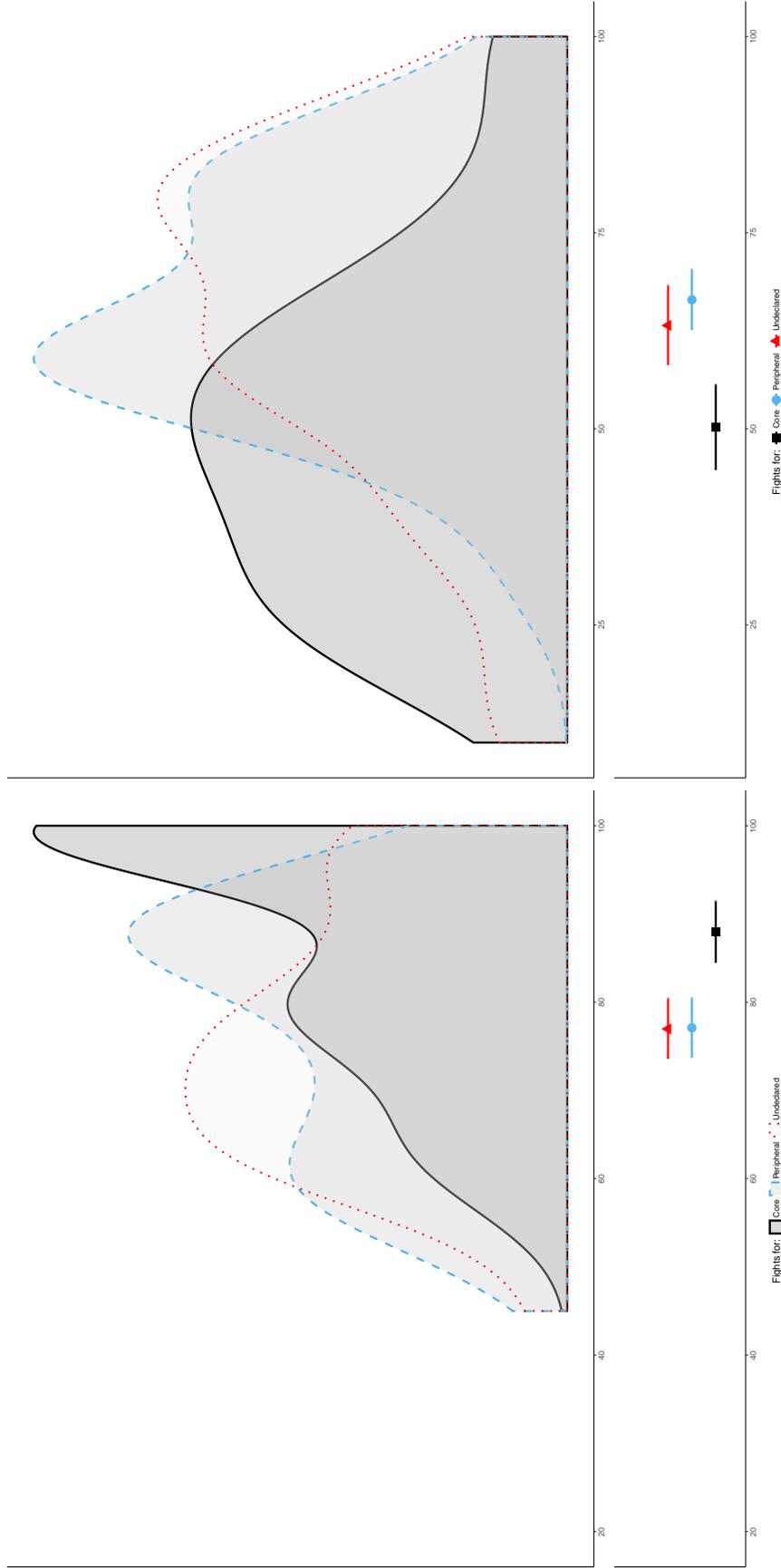
I visually represent the results in Figure 6. Panel A and B correspond to question **A** and **B** respectively. The x-axis reflects a respondent’s score to the question. I break-out responses into the three groups I make predictions about. The defender’s assessment when the challenger: (1) fights for core interests; (2) fights for peripheral interests; or (3) fights when core/peripheral interests are undefined. Each panel plots the density of responses for these groups. Below the density plots are coefficient plots that summarize the mean and 95% confidence intervals of each treatment group. Every hypotheses predicted that the core interest group (marked with a black square and solid line) would answer differently than the other two groups.

Starting with Panel A, I confirm my expectations about how defenders update beliefs about whether or not the challenger is willing to fight for core interests. Subjects that observed Bandaria fight for a declared core interest were more confident that Bandaria is resolved to fight for core interests than those that observed Bandaria fight for a peripheral interests, or those who had no information about Bandaria’s core interests. The permutation test confirms these differences are statistically significant at the 0.01 threshold. In contrast, there is no difference between those that observed Bandaria fight for peripheral interests and those who had no information about Bandaria’s core interests.

Turning to Panel B, I confirm my expectations about how defenders update beliefs about

Figure 6: Beliefs about intentions and resolve given different histories of fighting

(a) Response to **A**: Is Bandaria resolved to fight for core interests? (b) Responses to **B**: Will Bandaria fight opportunistically?



Responses are measured in % pr. True. They are broken down into 3 treatment groups and the pre-treatment condition. Distribution of responses plotted above. Mean with 95% confidence plotted below.

the challenger's long-term strategic intentions. Subjects that observed Bandaria fight for a core interests were less confident that Bandaria was resolved to fight opportunistically than those that observed Bandaria fight for a peripheral interest, or those that observed Bandaria fight in the counter-factual world where Bandaria's core interests were not clearly defined. The permutation test confirms these differences are statistically significant at the 0.01 threshold.

## 4.1 Further Analysis

In Appendix D.7 I complete all supplementary analyses reported in the pre-registration. I break down the results by the different diplomatic justifications. I show that if I constrain the analysis to either the ethnic or security justification that the result is robust. I use a variety of regression analyses to show that design features, including survey duration, sampling method, treatment assignment do not effect the result. I then used regression to explore if biographical features influenced the effect of treatment. I found that the treatment was robust when covariates were included and the results are consistent with my main findings.

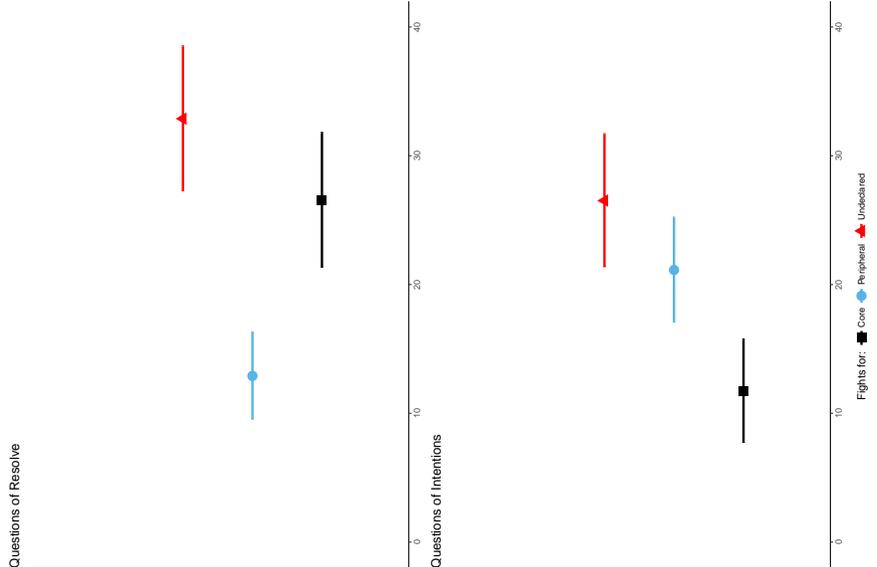
To increase my confidence that the results are not driven by a few outliers who update a lot, or confounding effects, I also analyze changes in individual level responses across different phases of the experiment. The distributions largely conform to my expectations about how subjects in each treatment group should update.

## 5 Discussion

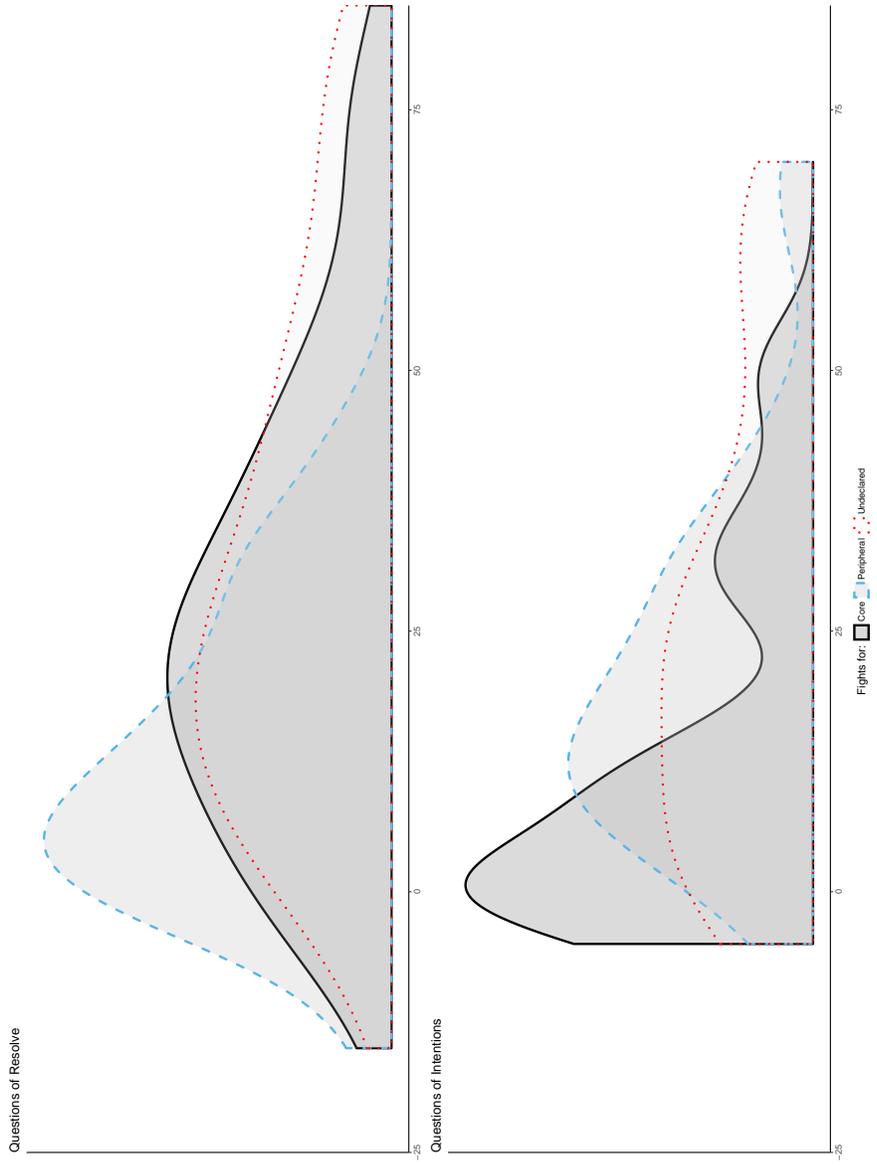
I developed a framework to understand the conditions under which states cultivate reputations for resolve and limited intentions. I argued that the context of what states are fighting over matters. In particular, states that fight over their core interests generate a reputation for resolve to fight for core interests in the future. States that fight for their peripheral interests cultivate a reputation as greedy. When core interests are undefined it

Figure 7: Changing Responses Across Time

(a) 95% Confidence Intervals.



(b) Distribution of Responses.



Responses are measured in changes in % pr. True from the second to third phase of the experiment. They are broken down into 3 treatment groups. Distribution of responses plotted above. Mean with 95% confidence plotted below.

is difficult to cultivate a reputation from crisis behavior across crises; leading defenders to rationally expect the worst.

I provided causal evidence to support my framework that directly observes how elites process information and form beliefs about national security crises. The empirics import recent advances in elite survey experiments from medical, legal and business research. These techniques use in sample randomization on a realistic vignette to provide strong internal validity. However, they also leverage sampling and solicitation methods, biographical checks, and other features to increase my confidence that the results should generalize.

This line of research has vital implications for American policy-makers who care intensely about the signals they send from their crisis behavior, and the inferences they can draw from how their rivals make threats. Past scholars have suggested that inferences are impossible, or cannot differentiate between China's decision to fight for Taiwan and Russia's decision to fight in Syria. This project takes seriously the context in which crises emerge. It paves the way for academics and policy-makers to integrate strategic dynamics and situational factors such as long-standing claims and the nature of the crisis.

## References

- Allison, G. T. and P. Zelikow (1971). *Essence of decision: Explaining the Cuban Missile Crisis*.
- Arber, S., J. McKinlay, A. Adams, L. Marceau, C. Link, and A. O'Donnell (2004, sep). Influence of patient characteristics on doctors' questioning and lifestyle advice for coronary heart disease: a UK/US video experiment. *The British journal of general practice : the journal of the Royal College of General Practitioners* 54(506), 673–8.
- Arber, S., J. McKinlay, A. Adams, L. Marceau, C. Link, and A. O'Donnell (2006, jan). Patient characteristics and inequalities in doctors' diagnostic and management strategies relating to CHD: A video-simulation experiment. *Social Science & Medicine* 62(1), 103–115.
- Barnett, C. (1986). *The collapse of British power*. Humanities Press International.
- Bullock, A. (1971). Hitler and the Origins of the Second World War. In *The Origins of the Second World War*, pp. 189–224. London: Macmillan Education UK.
- Camerer, C. F., G. Loewenstein, and M. Rabin (2004). *Advances in Behavioral Economics*. Princeton University Press.
- Carter, D. B. and H. E. Goemans (2011, apr). The Making of the Territorial Order: New Borders and the Emergence of Interstate Conflict. *International Organization* 65(02), 275–309.
- Cen, W. and J. A. Doukas (2017, oct). CEO personal investment decisions and firm risk. *European Financial Management* 23(5), 920–950.
- Collett, J. L. and E. Childs (2011). Minding the gap: Meaning, affect, and the potential shortcomings of vignettes. *Social Science Research* 40, 513–522.
- Dafoe, A., J. Renshon, and P. Huth (2014, may). Reputation and Status as Motives for War. *Annual Review of Political Science* 17(1), 371–393.
- Dafoe, A., B. Zhang, and D. Caughey (2016). Confounding in Survey Experiments.
- Danilovic, V. (2001, jun). The Sources of Threat Credibility in Extended Deterrence. *Journal of Conflict Resolution* 45(3), 341–369.
- Eysenbach, G. (2004). Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys. *Journal of medical Internet research* 6(3), e34.
- Fearon, J. D. (1994a, sep). Domestic Political Audiences and the Escalation of International Disputes. *American Political Science Review* 88(3), 577–592.
- Fearon, J. D. (1994b, jun). Signaling Versus the Balance of Power and Interests: An Empirical Test of a Crisis Bargaining Model. *Journal of Conflict Resolution* 38(2), 236–269.

- Fearon, J. D. (1995, jun). Rationalist Explanations for War. *International Organization* 49(03), 379.
- Fearon, J. D. (1997, feb). Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs. *Journal of Conflict Resolution* 41(1), 68–90.
- Feldman, H. A., J. B. McKinlay, D. A. Potter, K. M. Freund, R. B. Burns, M. A. Moskowitz, and L. E. Kasten (1997, aug). Nonmedical influences on medical decision making: an experimental technique using videotapes, factorial design, and survey sampling. *Health services research* 32(3), 343–66.
- Friedman, J., J. Lerner, and R. Zeckhauser. Behavioral Consequences of Probabilistic Precision: Experimental Evidence from National Security Professionals. *International Organization*.
- Fudenberg, D. (2006, sep). Advancing Beyond Advances in Behavioral Economics. *Journal of Economic Literature* 44(3), 694–711.
- Gilbert, M. (1972). *The Roots of Appeasement*. London: Weidenfeld and Nicolson.
- Gilpin, R. (1983). *War and Change in World Politics*, Volume 1983. Cambridge: Cambridge University Press.
- Glaser, C. L. (2010). *Rational Theory of International Politics*. Princeton: Princeton University Press.
- Gottfried, M. S. and R. F. Trager (2016, jun). A Preference for War: How Fairness and Rhetoric Influence Leadership Incentives in Crises. *International Studies Quarterly* 60(2), 243–257.
- Hafner-Burton, E. M., B. L. LeVeck, and D. G. Victor (2016, jan). How Activists Perceive the Utility of International Law. *The Journal of Politics* 78(1), 167–180.
- Hillgruber, A. (1974). England’s Place in Hitler’s Plans for World Domination. *Journal of Contemporary History* 9(1), 5–72.
- Hopf, T. (1994). *Peripheral visions : deterrence theory and American foreign policy in the Third World, 1965-1990*. University of Michigan Press.
- Huth, P. and B. Russett (1990, jun). Testing Deterrence Theory: Rigor Makes a Difference. *World Politics* 42(04), 466–501.
- Huth, P. K. (1997, sep). Reputations and deterrence: A theoretical and empirical assessment. *Security Studies* 7(1), 72–99.
- Jackson, M. O. and M. Morelli (2011). The Reasons for Wars: An Updated Survey. *The Handbook on the Political Economy of War*, 34–57.
- Jervis, R. (1978, jan). Cooperation Under the Security Dilemma. *World Politics* 30(02), 167–214.

- Joseph, M. (2017a). A Little Bit of Cheap-Talk is a Dangerous Thing.
- Joseph, M. (2017b). More Talk Less Action.
- Kertzer, J. D. (2016). *Resolve in international politics*. Princeton University Press.
- Kertzer, J. D. (2017, apr). Resolve, Time, and Risk. *International Organization* 71(S1), S109–S136.
- Kertzer, J. D. and K. M. McGraw (2012, jun). Folk Realism: Testing the Microfoundations of Realism in Ordinary Citizens. *International Studies Quarterly* 56(2), 245–258.
- Kitchen, M. (1986). *British policy towards the Soviet Union during the Second World War*. St. Martin's Press.
- Kydd, A. H. (2005). *Trust and Mistrust in International Relations*. Princeton, N.J.; Woodstock: Princeton University Press.
- Levendusky, M. S. and M. C. Horowitz (2012, mar). When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs. *The Journal of Politics* 74(02), 323–338.
- Levy, J. S. (2012, jul). Coercive Threats, Audience Costs, and Case Studies. *Security Studies* 21(3), 383–390.
- Levy, J. S., M. K. McKoy, P. Poast, and G. P. Wallace (2015, oct). Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory. *American Journal of Political Science* 59(4), 988–1001.
- Lieb, R. and B. Schwarz (2001, jan). The Year 2001 Survey: Ceo Perspectives On the Current Status and Future Prospects Of the Third Party Logistics Industry In the United States. *Supply Chain Forum: An International Journal* 2(2), 36–44.
- Ludbrook, J. and H. Dudley (1998, may). Why Permutation Tests Are Superior to t and F Tests in Biomedical Research. *The American Statistician* 52(2), 127.
- Lupton, R. and W. G. Jacoby (2016). The Reliability of the ANES Feeling Thermometers: An Optimistic Assessment. In *Annual Meetings of the Southern Political Science Association*.
- McManus, R. W. (2014, nov). Fighting words. *Journal of Peace Research* 51(6), 726–740.
- Mearsheimer, J. J. (2001). *The Tragedy of Great Power Politics*. New York: Norton.
- Mercer, J. (1996). *Reputation and International Politics*. Cornell University Press.
- Mintz, A. (2004, feb). Foreign Policy Decision Making in Familiar and Unfamiliar Settings. *Journal of Conflict Resolution* 48(1), 91–104.
- Moravcsik, A. (1998, oct). Taking Preferences Seriously: A Liberal Theory of International Politics: Erratum. *International Organization* 52(4), 229.

- Powell, R. (1999). *In the Shadow of Power: States and Strategies in International Politics*. Princeton, N.J: Princeton University Press.
- Press, D. G. (2007). *Calculating Credibility: How Leaders Assess Military Threats*. Cornell University Press.
- Rashad Abdel-Khalik, A. (2014, sep). CEO Risk Preference and Investing in R & D. *Abacus* 50(3), 245–278.
- Redding, R. E., M. Y. Floyd, and G. L. Hawk (2001, jul). What judges and lawyers think about the testimony of mental health experts: a survey of the courts and bar. *Behavioral Sciences & the Law* 19(4), 583–594.
- Renshon, J. (2015, may). Losing Face and Sinking Costs: Experimental Evidence on the Judgment of Political and Military Leaders. *International Organization* 69(03), 659–695.
- Renshon, J. (2016, aug). Status Deficits and War. *International Organization* 70(03), 513–550.
- Renshon, J., A. Dafoe, and P. Huth (2018, jan). Leader Influence and Reputation Formation in World Politics. *American Journal of Political Science*.
- Ripsman, N. M. and J. S. Levy (2008, oct). Wishful Thinking or Buying Time? The Logic of British Appeasement in the 1930s. *International Security* 33(2), 148–181.
- Rosato, S. (2015, jan). The Inscrutable Intentions of Great Powers. *International Security* 39(3), 48–88.
- Sartori, A. E. (2002, feb). The Might of the Pen: A Reputational Theory of Communication in International Disputes. *International Organization* 56(1), 121–149.
- Sartori, A. E. (2005). *Deterrence by diplomacy*. Princeton University Press.
- Saunders, E. (2011). *Leaders at war : how presidents shape military interventions*. Cornell University Press.
- Schultz, K. A. (2001a). *Democracy and coercive diplomacy*. Cambridge University Press.
- Schultz, K. A. (2001b, feb). Looking for Audience Costs. *Journal of Conflict Resolution* 45(1), 32–60.
- Slantchev, B. L. (2003, feb). The Power to Hurt: Costly Conflict with Completely Informed States. *American Political Science Review* 97(01), 123.
- Steiner, P. M., C. Atzmüller, and D. Su (2017, jun). Designing Valid and Reliable Vignette Experiments for Survey Research: A Case Study on the Fair Gender Income Gap. *Journal of Methods and Measurement in the Social Sciences* 7(2).
- Strasser, H. and C. Weber (1999). On the asymptotic theory of permutation statistics. Technical report, WU Vienna University of Economics and Business, Vienna.

- Streich, P. and J. S. Levy (2007, apr). Time Horizons, Discounting, and Intertemporal Choice. *Journal of Conflict Resolution* 51(2), 199–226.
- Tang, S. (2005, jan). Reputation, Cult of Reputation, and International Conflict. *Security Studies* 14(1), 34–62.
- Tingley, D. H. (2011, jun). The Dark Side of the Future: An Experimental Test of Commitment Problems in Bargaining1. *International Studies Quarterly* 55(2), 521–544.
- Tingley, D. H. and B. F. Walter (2011, aug). Can Cheap Talk Deter?: An Experimental Analysis. *Journal of Conflict Resolution* 55(6), 996–1020.
- Tomz, M. (2007). *Reputation and International Cooperation: Sovereign Debt Across Three Centuries*. Princeton University Press.
- Trager, R. F. (2011, jul). Multidimensional Diplomacy. *International Organization* 65(03), 469–506.
- Uzonyi, G., M. Souva, and S. N. Golder (2012, dec). Domestic Institutions and Credible Signals. *International Studies Quarterly* 56(4), 765–776.
- Waltz, K. N. (1979). *Theory of international politics*. Reading, Mass: McGraw-Hill.
- Weeks, J. L. (2008). Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization* 62(01), 35–64.
- Weinberg, G. L. (1980). *The foreign policy of Hitler's Germany : starting World War II, 1937-1939*. University of Chicago Press.
- Weisiger, A. and K. Yarhi-Milo (2015, may). Revisiting Reputation: How Past Actions Matter in International Politics. *International Organization* 69(2), 473–495.
- Wood, B. D. (2012). *Presidential saber rattling : causes and consequences*. Cambridge University Press.
- Yarhi-Milo, K. (2014). *Knowing the adversary : leaders, intelligence, and assessment of intentions in international relations*. Princeton University Press.
- Yoder, B. K. (2019, jan). Retrenchment as a Screening Mechanism: Power Shifts, Strategic Withdrawal, and Credible Signals. *American Journal of Political Science* 63(1), 130–145.

## **A The mediating effect of core interests**

Please contact author for detailed proof.

## **B Searching for (and finding) Press's, equilibrium**

Please contact author for detailed proof.

## C Survey Text

Below is the text that subject see in the experiment. Each section corresponds with a screen in the experiment.

### Screen 1: Introduction

This exercise simulates an intelligence assessment. The scenario is fictitious and not designed to resemble any particular country. However, all the information draws from war game scenarios that senior decision-makers participate in.

By taking this exercise seriously you will improve how we, as policy-makers and intelligence professionals, estimate the intentions of other countries.

### Screen 2: Prompt

#### The Task

The country of Bandaria is experiencing unprecedented economic growth. We know very little about Bandaria's long-term foreign policy goals.

Later this year, the U.S. President will meet the Bandarian Prime Minister. Before that meeting, the president needs information about Bandaria's foreign policy interests and their willingness to use force.

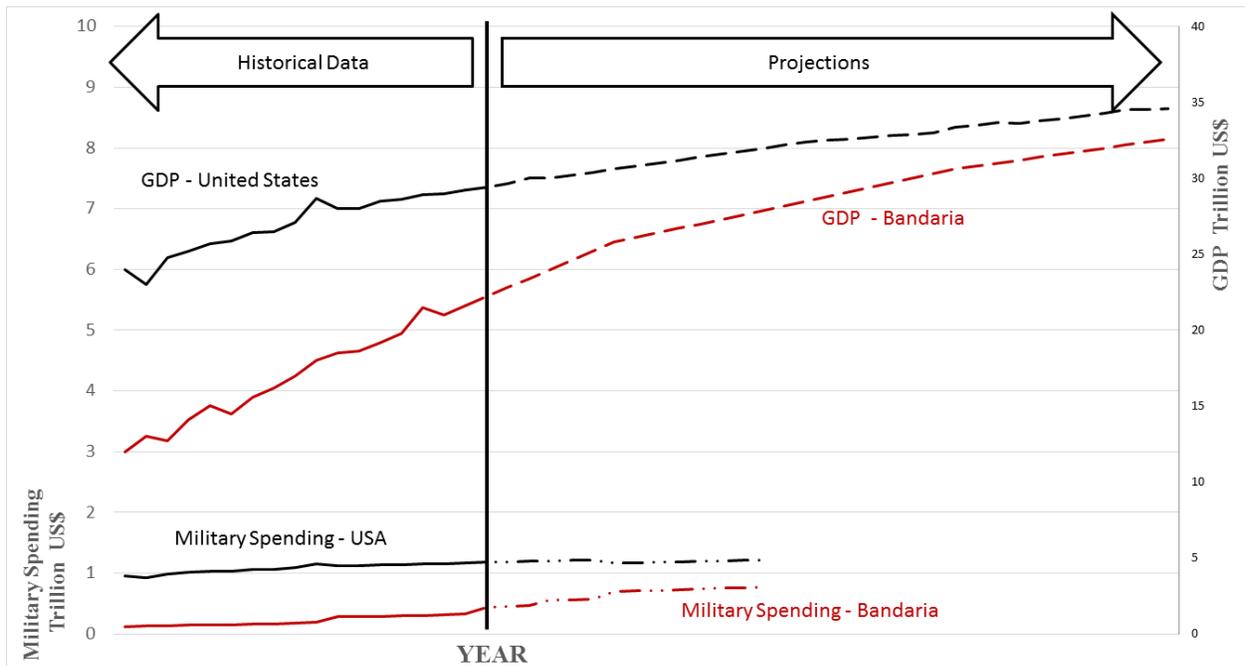
Evaluate the following information and provide an assessment of Bandaria's long-term intentions.

### Slide 3: Baseline Vignette

#### Economic Trends

In the next 50 years, Bandaria will become one of the largest economies in the world. Bandaria's projected growth is compared to the United States' in the figure below.

Figure 8: Comparing the United States to Bandaria on GDP Growth and Military Expenditure.



### Military Trends

Last year, Bandaria announced it will spend 2.6% of GDP on military modernization over the next decade. If Bandaria increases military spending as they project, they will be the largest military power on their continent in 10 years, although still much weaker than the United States.

Experts disagree about the trajectory of Bandaria’s military growth beyond 10 years. Some think Bandaria has aggressive long-term intentions. They note that Bandaria has the resources to increase military spending long into the future. Some think Bandaria has benign long-term intentions. They note Bandaria would prefer to spend their surplus on domestic programs after an initial military modernization effort.

## Political Trends

Experts do not regard Bandaria as democratic. Bandarians elect a Prime Minister through competitive single-party elections. However, the Prime Minister shares power with the Bandarian King.

The government is popular. The Prime Minister has stamped out local corruption and put in place welfare reforms to lift the poorest citizens out of poverty. 90% of Bandarians belong to the same ethnic and religious group. The minorities are well integrated into society. Regional experts agree that the system of government will remain stable over the next thirty years.

Figure 9: Regional Map



## Surrounding Countries

Bandaria is situated on a continent with three other countries: Arcadia, Lakesh and New Kasper.

There are no significant natural resource deposits on the continent.

## **Trade and Diplomacy**

Bandaria exchanges diplomats with all its regional neighbors as well as many other countries. It is a member of the United Nations and other international organizations.

Bandaria benefits from strong trade ties with its regional neighbors, the United States and many other countries.

## **Possible Strategic Tensions**

- One province of Arcadia, called Greywall, is populated mainly by ethnic Bandarians. Ethnic Bandarians complain that they are treated as second class citizens in Arcadia. Ethnic Bandarians in Greywall are, on average, of lower socio-economic status than native Arcadians. There is no concentration of ethnic Bandarians elsewhere in the region.
- The Lebang Bay contains the only deep water ports on the continent. It is the major trade thoroughfare and strategic choke point in the region. New Kasper controls the Lebang Bay and governs the territory that surrounds it. 50 years ago, New Kasper and Bandaria disputed fishing rights off their coast. During that dispute, New Kasper blocked Bandarian trade through the Lebang Bay for one week. The effect on the Bandarian economy was enormous. Bandaria and New Kasper have resolved their disputes and now have normal trade and diplomatic relations. Bandaria faces no other key security vulnerabilities elsewhere in the region.

## Screen 4: Baseline Response

### C.0.1 Assessment Based On Available Information

Military analysts warn that Bandaria will soon be powerful enough to revise the international status-quo if it wants to. The question is: does it want to?

Based on the information you've seen, provide an assessment of Bandaria's intentions.

You will be asked to write your assessment in your own words and answer some standardized questions.

Note: A summary of what you've read so far is below, but you can review all the information by clicking here.

**[Question]** What is your assessment of Bandaria's long-term intentions?

[Text Box]

**[Prompt]** What is the percentage probability that the following statements are true?

Move the slider towards certainly true (100%) or certainly false (0%) based on how confident you are that the statement is true.

- No matter what territorial concessions are made Bandaria will demand more if it is given the opportunity to do so.
- Bandaria will use military force to achieve its core foreign policy objectives.
- Bandaria will use military force to expand its borders whenever the opportunity presents itself.
- Although there are many military objectives that Bandaria might pursue, a single target stands out as the most likely.

[slide rules 0-100 that move in 5 point increments. Labeled Certainly False, Certainly True at endpoints]

[**Question**] In the last question you were asked to think about a most likely target. Click on the map where that most likely target is.

[Click-Map of Bandaria]

[**Question**] Once Bandaria achieves this objective, it will stop making demands.

## **Screen 5: Treatment 1 + Response 2**

### **C.0.2 Notes**

Subjects are randomly assigned into 1 of 3 treatments. They are then all subject to the same standardized questions that appear below. The treatments read as follows.

### **C.0.3 Minutes of Diplomatic Meeting**

### **C.0.4 Security Treatment**

In a private meeting, the American president asked the Prime Minister of Bandaria to explain Bandaria's military spending. The Prime Minister replied:

We worry about our security. We have key vulnerabilities in the Lebang Bay. Although our relations with New Kasper are good today, anything can happen tomorrow. Concerns for our security motivate our military policy.

He then said,

Of course we want to resolve this issue peacefully. But New Kasper does not realize just how concerned we are about our international security. We will use any means necessary to assure that our security is protected from potential foreign influence.

He continued,

Once our international security is assured, we have no reason to expand our military. All of our other foreign policy and regional concerns are less important and can be managed through UN participation, diplomacy and negotiation.

Experts note that security concerns in the Lebang Bay have been central to Bandarian foreign policy over the past 10 years. Bandarian elites referred to security issues in private diplomatic conversations and public speeches consistently over the past 10 years.

### **C.0.5 Ethnic Treatment**

In a private meeting, the American president asked the Prime Minister of Bandaria to explain Bandaria's military spending. The Prime Minister replied:

Grave injustices have been done to ethnic Bandarians. We have a long history of supporting our Bandarian brothers in Arcadia. Ethnic-national concerns motivate our military policy.

He then said,

Of course we want to resolve this issue peacefully. But Arcadia does not realize just how concerned we are about our ethnic kin. We will use any means necessary to ensure our ethnic kin are well governed.

He continued,

Once our ethno-nationalist goals are assured, we have no reason to expand our military. All of our other foreign policy and regional concerns are less important and can be managed through UN participation, diplomacy and negotiation.

Experts note that ethnic nationalism concerns in Greywall have been central to Bandarian foreign policy over the past 10 years. Bandarian elites referred to ethnic-nationalism in private diplomatic conversations and public speeches consistently over the past 10 years.

## C.0.6 Control

In a private meeting, the American president and the Prime Minister of Bandaria exchanged sentiments of mutual respect. Both called for increased confidence building measures.

## C.0.7 Task

Given what you now know, we will ask you some more questions about Bandaria's long-term intentions. Some questions are the same, others are new.

Note: A summary is below, but you can review all the information you've seen before by clicking [here](#).

**[Prompt]** What is the percentage probability that the following statements are true?

Move the slider towards certainly true (100%) or certainly false (0%) based on how confident you are that the statement is true.

- No matter what territorial concessions are made Bandaria will demand more if it is given the opportunity to do so.
- Bandaria will use military force to achieve its core foreign policy objectives.
- Bandaria will use military force to expand its borders whenever the opportunity presents itself.
- We can trust what the Bandarian Prime Minister said about Bandaria
- Although there are many military objectives that Bandaria might pursue, a single target stands out as the most likely.

[slide rules 0-100 that move in 5 point increments. Labeled Certainly False, Certainly True at endpoints]

[**Question**] In the last question you were asked to think about a most likely target. Click on the map where that most likely target is.

[Click-Map of Bandaria]

[**Question**] Once Bandaria achieves this objective, it will stop making demands.

[**Question**] What is your assessment of Bandaria's long-term intentions now? [text]

## **Screen 6: Treatment 2 + Response 3**

### **C.0.8 Notes**

Subjects are randomly assigned into 1 of 2 treatments. They are then all subject to the same standardized questions that appear below. The treatments read as follows.

### **C.0.9 Breaking news**

### **C.0.10 Security Treatment**

The Bandarian military occupied territory surrounding the Lebang Bay in New Kasper. The Bay is the primary security issue in the region. The Bandarian Prime Minister announced plans to annex the Bay but promised fair treatment and reparations for aggrieved New Kasper citizens and businesses.

The Bandarian Prime Minister insists that these events are entirely consistent with Bandaria's interests long known to the rest of the world and Bandaria remains committed to peace and stability generally. He pledged to respect the sovereignty of all other countries in the region.

### C.0.11 Ethnic Treatment

The Bandarian military occupied Greywall in Arcadia. Greywall is populated by ethnic Bandarians. The Bandarian Prime Minister announced plans to annex Greywall but promised fair treatment and reparations for aggrieved Arcadian citizens and businesses.

The Bandarian Prime Minister insists that these events are entirely consistent with Bandaria's interests long known to the rest of the world and Bandaria remains committed to peace and stability generally. He pledged to respect the sovereignty of all other countries in the region.

### C.0.12 Task

Given what you now know, we will ask you some more questions about Bandaria's long-term intentions. Some questions are the same, others are new.

Note: A summary is below, but you can review all the information you've seen before by clicking [here](#).

**[Prompt]** What is the percentage probability that the following statements are true?

Move the slider towards certainly true (100%) or certainly false (0%) based on how confident you are that the statement is true.

- No matter what territorial concessions are made Bandaria will demand more if it is given the opportunity to do so.
- Bandaria will use military force to achieve its core foreign policy objectives.
- Bandaria will use military force to expand its borders whenever the opportunity presents itself.
- We can trust what the Bandarian Prime Minister said about Bandaria

- Although there are many military objectives that Bandaria might pursue, a single target stands out as the most likely.

[slide rules 0-100 that move in 5 point increments. Labeled Certainly False, Certainly True at endpoints]

[**Question**] In the last question you were asked to think about a most likely target. Click on the map where that most likely target is.

[Click-Map of Bandaria]

[**Question**] Once Bandaria achieves this objective, it will stop making demands.

[**Question**] Does this information change your assessment of Bandaria's long-term intentions? If so, why? If not, why not? [text]

## **Screen 7-9: Manipulation Checks, Biographic Questions, Post-Treatment Questionnaire**

Subjects are given a battery of questions that determine (1) Their eligibility in the sample; and (2) if they paid attention to the questions.

Subjects are asked a series of questions about their work function.

Subjects are randomly assigned into 4 groups and asked follow up questions about how they use information in their work.

## D Supplementary Analysis

Below I provide additional information about the survey experiment. Appendix B presented the text of the survey. This appendix presents summary statistics, additional analyses and also information about sampling procedures.

### D.1 Sampling Procedure

In this section I explain the sampling procedure. I describe my method of solicitation, then two types of checks I used to rule out inappropriate subjects: sample inclusion questions (that ensure subjects are elites); and attention checks (that ensure subjects properly read the questions).

#### D.1.1 Two distinct samples:

The sampling method relied on a convenience sample of policy elites. I used two distinct sampling methods that targeted different groups of elites. I provided these groups with different links to two identical surveys. First, I sent solicitation emails through institutions that interact with mid-level foreign professionals. Three Master’s Degree Programs that accept only mid-Career American Foreign Policy and Intelligence Professionals sent solicitation emails to their alumni network. The Australian Embassy in the United States and the American Embassy in New Zealand solicited their policy, defense and analytical staff (approximately 800 solicitations).

Second, I used an elite “snowball” sampling technique. During field research, I developed contacts within 31 foreign policy professionals. I asked them to distribute solicitations through their professional networks within my sample frame.<sup>28</sup> Employers include the State Department, Office of the Secretary of Defense, various military and civilian intelligence

---

<sup>28</sup>To be clear, I did not ask the people I knew directly to take the survey, only to distribute it through their network.

agencies and staff for the Armed Services Congressional Committees. In the snowball sample, I asked my professional network to extend a solicitation email to subjects that they believed fit my sampling frame.

Table 4 reports the results of two linear regression analysis. The dependent variables in columns (a) and (b) are the responses to questions **A** and **B** respectively. In the analysis include the treatment — whether or not the subject observed Bandaria fight for a core or peripheral interest (the control is omitted). We also included a control for sampling method. The table cofirms that the treatment significantly predicts how subjects respond, but the method of treatment does not.

Table 4: Sampling Method

	Response to Strategic Intentions (1)	Response to Resolve (2)
Fought for Core Interests	10.426** (3.759)	-16.880** (5.100)
Institutional Sample (Y/N)	-2.354 (3.834)	-3.465 (5.202)
Constant	78.236** (3.171)	68.128** (4.303)
Observations	60	60

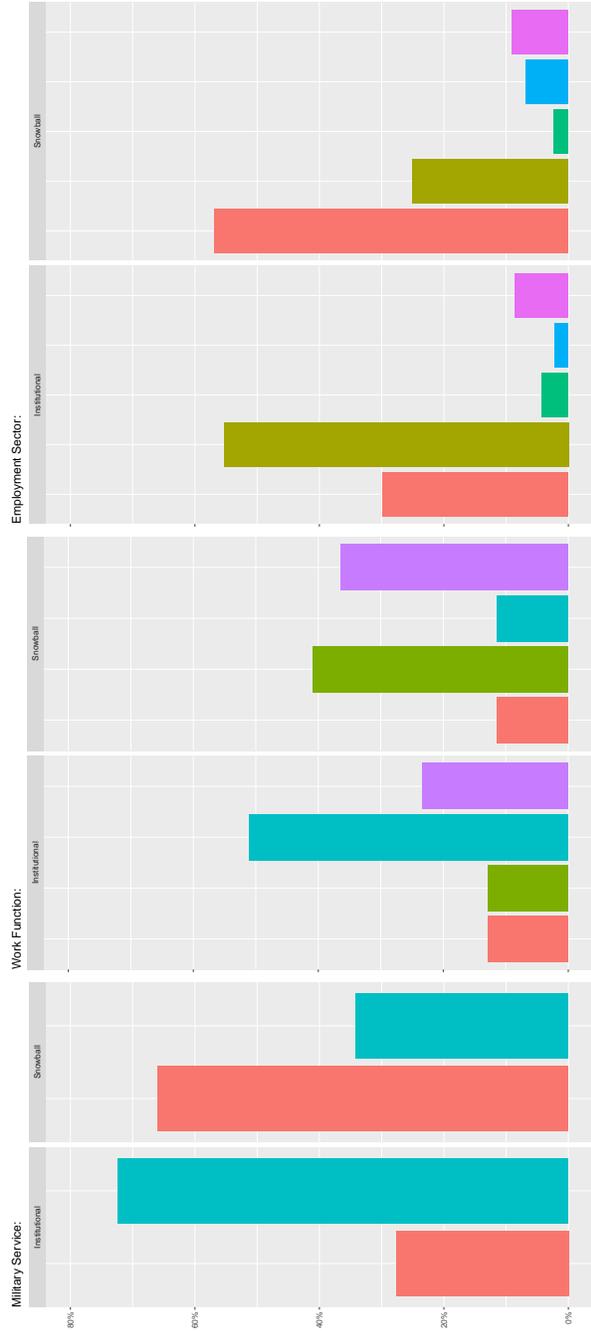
*Note:*

\*p<0.05; \*\*p<0.01

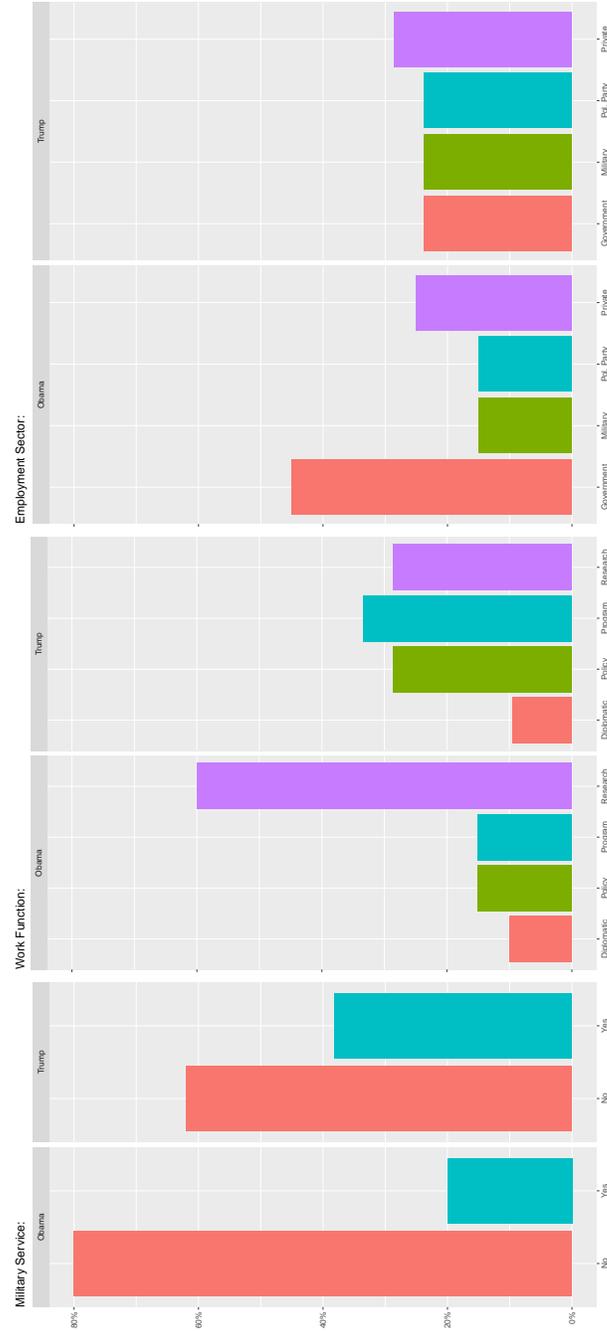
Figure 10(a) summarizes the biographical information from the two sub-samples. The snowball sample had considerably less military experience then the institutional sample, and worked in civilian government agencies more frequently. Figure 10(b), reports biographical information from the NSC principle members selected by president Obama, and president Trump. The comparisons demonstrate that the variation across my sampling methods corresponds with actual variation in NSC selection.

Figure 10: Biographical information:

(a) Survey participants by sampling method.



(b) NSC principals by president.



Each plot represents the percentage of responses to a single biographical question. Within each plot, there are two panels that correspond to sub-groups.

### D.1.2 Solicitation Email

The following text is the solicitation email sent to subjects from the snowball sample:<sup>29</sup>

Please take part in a study that simulates a foreign policy assessment. We need foreign policy, and defense professionals, broadly defined, to participate. We are contacting you through a friend or colleague that knows about the project and recommended you as an ideal participant.

To let you know about the survey:

1. The research simulates a foreign policy assessment. The information is fictitious, but drawn from simulations that very senior leaders have participated in.
2. It is totally anonymous. We collect no meta-data or identifying information. We are surveying several organizations and all answers are pooled. Thus, we do not know who responds or even which organization they work for. An external survey firm (Qualtrics) guarantees the anonymity of the results.
3. It takes about 20 minutes. Pilot subjects took between 15-25 minutes to complete. But all noted it took their full attention for that time.
4. The survey will end on June 29th.
5. Take it at your convenience on any desktop computer (Smartphones are disabled): Just click the link below and the survey will begin. We disable the survey on mobile devices to make sure we can protect your identity.
6. Its fun. Pilot subjects really enjoyed taking it.

Please start the survey when you have 20 minutes to focus on it. To begin, click the link below:

[LINK]

---

<sup>29</sup>The Solicitation for the institutions is similar, and therefore omitted.

As you can see, the link directs you to [Institution Name]’s Qualtrics research account. [Institution Name] is a research university in [City]. Qualtrics is the leading survey research firm worldwide. Qualtrics makes sure the information is anonymous.

Your participation will make a meaningful difference. We really appreciate your help!

For more information please contact: [My Email]. We’d be happy to tell you more about the research.

Cordially,

[My Name - removed for review]

## D.2 Response Rates

217 subjects clicked on the link and read the prompt. 198 subjects clicked past the prompt page to read the baseline vignette. 138 subjects clicked to the next page and answered at least one question. 131 subjects read the two treatments and answered all the questions. Since my analyses focuses on answers in the last phase of the experiment I could only analyze subjects that made it to the end of the experiment.

There is a noticeable drop-off between those that observed the baseline vignette and those that clicked to the next screen and answered one question. The probable reason is that the baseline vignette is approximately three pages long and contains a great deal of information. Many subjects probably observed this information and chose to end participation.

Fortunately, I can be certain that differential treatment effects did not cause this large drop-off in participation because all subjects receive the same baseline information. Thus, the drop-off occurred before subjects received different information.

The attrition rate between those that answered one question (before differential treatment effects) and those that completed the survey (after treatments) is just 7 subjects. As Table 5 shows, these subjects are dispersed across treatment groups.

	Ethnic Int.	Security Int.
Control Message	0	2
Ethic Message	2	0
Security Message	1	2

Table 5: Dropout Rates Between Answering First Question and All Questions.

### D.3 Sample Inclusion Questions

Subjects were asked 2 questions to determine if they fit the sample frame:

1. What is the position of the highest ranking government official you have briefed during your career? [Text]
2. Do you work on foreign policy issues? [Y/N]

Subjects were ruled out if they answers to these answers to these questions indicated they were outside the sampling frame. Additionally, 7 subjects completed the survey and passed all attention checks but did not answer any of the biographical questions including these sample inclusion questions. It is possible, that these subjects did not want to provide personal information on an online survey platform because of their work affiliation. Thus, these subjects may be appropriate for inclusion. Nevertheless, I chose to omit these subjects from the sample for two reasons: (1) I wanted to be as sure as possible that the sample was elite; (2) smoe property of these subjects that led them to be reluctant to complete biographical questions may have also interacted with the treatment group. By eliminating them, I can omit factors that may have effected sample heterogeneity.

### D.4 Did the subjects pay attention

The experiment contains three checks to ensure that subjects paid attention and that attention did not vary with treatment. First, I included attention checks. Second, I recorded long-form text responses and timed the survey to see if subjects spent a reasonable amount of time on the survey. Third, I then re-ran the survey with Mechanical Turk workers to

compare the attention statistics of elites to mechanical turk workers: 73% of elite respondents correctly answered these attention checks (twice as many as M-Turk workers). The elites took on average 29 minutes (50% longer than the M-Turk Workers) and had an average open text response of 509 characters (3 times longer than the average M-Turk worker).

#### D.4.1 Attention Checks

Subjects were asked two post-treatment questions to test if they read the information in the vignette. Of those that passed sample inclusion questions,<sup>30</sup> and completed the survey, 32 failed at least one of the attention checks. Table 6 shows that they are well dispersed across treatment groups.

	Passed	Failed
Core interest crisis	29	11
Peripheral interest crisis	31	13
Interests Undefined	33	8

Table 6: Elites Who Passed Attention Checks?

#### D.4.2 Meta-data

Table 7 reports the time it took subjects to respond and the length of their text responses.

Table 7: Attention Data By Treatment

Treatment Group	Median Minutes
Core interest crisis	25
Peripheral interest crisis	36
Interests Undefined	29
Full Sample	33

Treatment Group	Median Character Length
Core interest crisis	457
Peripheral interest crisis	623
Interests Undefined	498
Full Sample	509

<sup>30</sup>i.e subjects I am confident are elites.

Interestingly, the subjects that observed a peripheral crisis took longer to respond than the other groups. In other settings, subjects who observe unusual or out of place information feel anxious and spend more time evaluating information before they reach their assessment. It is consistent with my theory that observing the peripheral interest treatment would behave this way.

## D.5 Balance tests in Summary Statistics

The following tables breaks out treatment groups by various covariates.<sup>31</sup> I broke out the tables to show that there is a good dispersion of covariates across the sample and most subjects responded to biographical information if they completed the survey.

Subjects also cover a broad range of work functions and organizational affiliations. Notably, the NSC includes advisors from defense, treasury, commerce, the intelligence community, state department and so on. It includes analysts as well as operations staff who deal with more practical matters. Thus, the broad scope of subjects' experiences captures a certain amount of diversity that one might find on the NSC.<sup>32</sup> Further, two subjects briefed a Head of State on a foreign policy issue. 30% had briefed a member of the NSC. An additional 52% had briefed an Ambassador, Member of Congress or the Senate, or a General.

---

<sup>31</sup>Note all subjects answered all of the biographical questions. Thus, the numbers do not correspond. I restrict my attention to the sample I analyze in the paper.

<sup>32</sup>Not all subjects answered all biographical information. The numbers therefore do not always add up to 93.

Table 8: Biographical Data By Treatment

	Core	Undefined	Peripheral	Total
Yes	24	21	22	67
No	5	12	9	26
Total	29	33	31	93

	Core	Undefined	Peripheral	Total
Yes	14	20	15	49
No	15	12	15	42
Total	29	32	30	91

	Core	Undefined	Peripheral	Total
Civil Government Agency	15	12	12	39
Military Agency	9	15	13	37
Political Party	2	1	1	4
Private Sector	3	2	3	8
Other	0	2	1	3
Total	29	32	30	91

	Core	Undefined	Peripheral	Total
Research/Analysis	9	6	12	27
Policy-Making	13	7	4	24
Programmatic Work/Operations	2	14	13	29
Diplomacy/Political Communication	5	5	1	11
Total	29	32	30	91

	Core	Undefined	Peripheral	Total
Head of State	1	1	0	2
Cabinet Official/Chairman of Joint Chiefs	9	8	7	24
Amb./General/Senator/Congress	14	17	15	46
Other Elites	3	5	8	16
Total	27	31	30	88

Table 9: Sector By Military Service:

	Military Service	No Service
Civil Government Agency	10	29
Military Agency	33	4
Political Party	1	3
Private Sector	4	4
Other	1	2

## D.6 Dissociating Bandaria from real cases

I took several steps to make sure subjects did not associated Bandaria with a real case. I administered the survey towards the end of the American presidential primary when all eyes were on domestic politics. The major interstate incident during this period was Brexit—an event unrelated to military intervention. I chose a fictitious scenario with fake names and a map based on merging and manipulating American congressional districts. Finally, I told subjects that the scenario was designed not to resemble a particular case, but rather that the information was derived from war games that senior policymakers had participated in. In a mechanical turk pilot I provided test subjects parts of the vignette and asked them what pieces of information reminded them of particular cases. I then developed a fact base that was not similar to any single case.

To make sure these steps were effective, I asked subjects two questions after they completed the survey.

- Did you have a particular event, either current or historical, in mind at the beginning of this survey? If yes which one?
- Did you have a particular event, either current or historical, in mind at the end of this survey? If yes which one?

70% of subjects reported no event in their mind at the beginning and at the end. Of the 30% that had an event in mind at the beginning of the experiment, 3/4 reported that

they had a different case in mind at the end of the experiment. Further, there was enormous variation in what subjects identified. Subjects identified Vietnam, China, Afghanistan, Iraq, Russia, Iran and Sudan as cases. The case most frequently identified was Russia/USSR, and only 8% of participants identified it (it is unclear if they were describing Russia today or the Soviet Union).

## D.7 Regression Analysis

To further demonstrate the proper application of randomization, I report regression results that include covariates above. The sample only includes those that saw Bandaria fight for either core or peripheral interests. I omit the control group. Since my sample size is already small, I consider potential confounding effects separately.

The procedure is as follows. First, I subset the data to omit those that received a control. Next, I estimate regressions of the following form:

$$High\ resolve_i = \beta_0 + \beta_1 Core\ interest + \beta_k Control_j + \epsilon \quad (8)$$

$$Opportunistic_i = \beta_0 + \beta_1 Core\ interest + \beta_k Control_j + \epsilon \quad (9)$$

Where the outcome variable is the subject's response at the third stage of the experiment. The main independent variable is a binary indicator, equal to one if subject received consistent treatment, and 0 if they received an inconsistent treatment. The equation then includes some controls (indexed by j to make explicit I estimate different models for each set of controls). I consider the following controls:

1. Message Type: (1) Security Message, (2) Ethnic Message
2. Military Service
3. Citizenship: (1) American, (0) Australian/New Zealand

4. Highest Ranking Official: Ordinal variable
5. Employment Sector
6. Work Function
7. Survey Duration: continuous variable

Each table presents the results to both Dependent Variables with the same suite of controls.

The results clearly show that the treatment is reliably consistent with high confidence in 15 out of 16 models. The one model where treatment is no longer significant includes a categorical variable with 5 categories. Even that model is significant with 89% confidence.

Table 10: Type of Core Interests

	High Resolve	Strategic Intentions
	(1)	(2)
Fought for core interest	10.940** (3.726)	-16.674** (5.033)
Security Core Interests (Y/N)	1.110 (3.757)	-4.122 (5.076)
Observations	60	60

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

## D.8 Analysis across different phases of the experiment

I report changes in responses across different phases of the experiment in Figure 11. That is, I broke out responses into their treatment groups: consistent, inconsistent and counterfactual. I then took their third phase responses (after they observed Bandaria’s decision to fight) and subtracted them from their second phase responses (after they observed Bandaria’s declared core interests). This analysis shows how subjects update at an individual level. By analyzing individual level responses, I can increase my confidence that a few outlier

Table 11: Military Service

	High Resolve (1)	Strategic Intentions (2)
Fought for core interest	10.082** (3.691)	-16.690** (5.104)
Military Service (Y/N)	0.918 (3.691)	1.690 (5.104)
Observations	59	59

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

Table 12: Citizenship

	High Resolve (1)	Strategic Intentions (2)
Fought for core interest	11.691** (3.658)	-16.267** (5.088)
American Citizen (Y/N)	-7.266 (4.322)	-0.102 (6.012)
Observations	60	60

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

Table 13: Seniority

	High Resolve (1)	Strategic Intentions (2)
Fought for core interest	10.083* (3.805)	-16.046** (5.418)
Briefed NSC principle (Y/N)	-1.265 (2.612)	1.263 (3.720)
Observations	57	57

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

Table 14: Employment Sector

	High Resolve	Strategic Intentions
	(1)	(2)
Fought for core interest	9.102** (3.346)	-17.186** (5.262)
Military	0.138 (3.648)	-2.130 (5.738)
Other	-30.684* (12.951)	-18.251 (20.370)
Political Party	1.581 (7.669)	-10.127 (12.061)
Private Sector	-19.402** (5.684)	4.508 (8.939)
Observations	59	59

*Note:*

Diplomat is baseline category

*Note:*

\* p<0.05; \*\* p<0.01

Table 15: Work Function

	High Resolve	Strategic Intentions
	(1)	(2)
Fought for core interest	6.780 (4.262)	-19.171** (5.867)
Policy-making	6.909 (4.782)	-6.791 (6.584)
Programs/operations	0.954 (4.892)	-12.660 (6.735)
Diplomatic	10.708 (6.699)	0.260 (9.223)
Observations	59	59

*Note:* Baseline category is research  
*Note:* Core interest treatment is significant at 11%  
*Note:* \* p<0.05; \*\* p<0.01

Table 16: Duration

	High Resolve	Strategic Intentions
	(1)	(2)
Fought for core interest	10.512** (3.646)	-16.765** (4.928)
Duration	-3.383 (2.251)	-5.097 (3.042)
Observations	60	60

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

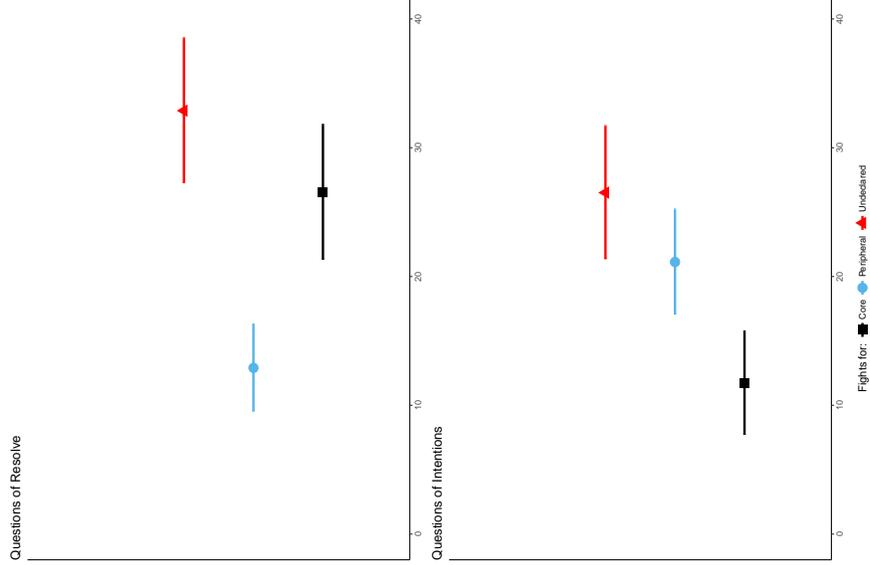
individuals are not driving the result in a small sample size and also show that individuals update as I expect them to across different phases of the experiment.

Panel A plots the mean and 95% confidence intervals broken out by the three treatment groups. The top row reports responses to **B**, the bottom row reports responses to **A**. I stacked responses so it is easy to compare how the same treatment group responds differently to different questions. Subjects that observed Bandaria fight for a core interest (black square) increased their confidence that Bandaria was highly resolved to fight for core interests more than they increased their confidence that Bandaria was willing to fight opportunistically. The opposite is true for subjects that observed Bandaria fight for a peripheral interest (blue dot). When Bandaria's core interests are not defined (the red triangle) subjects update their responses to both questions the same way following a military intervention.

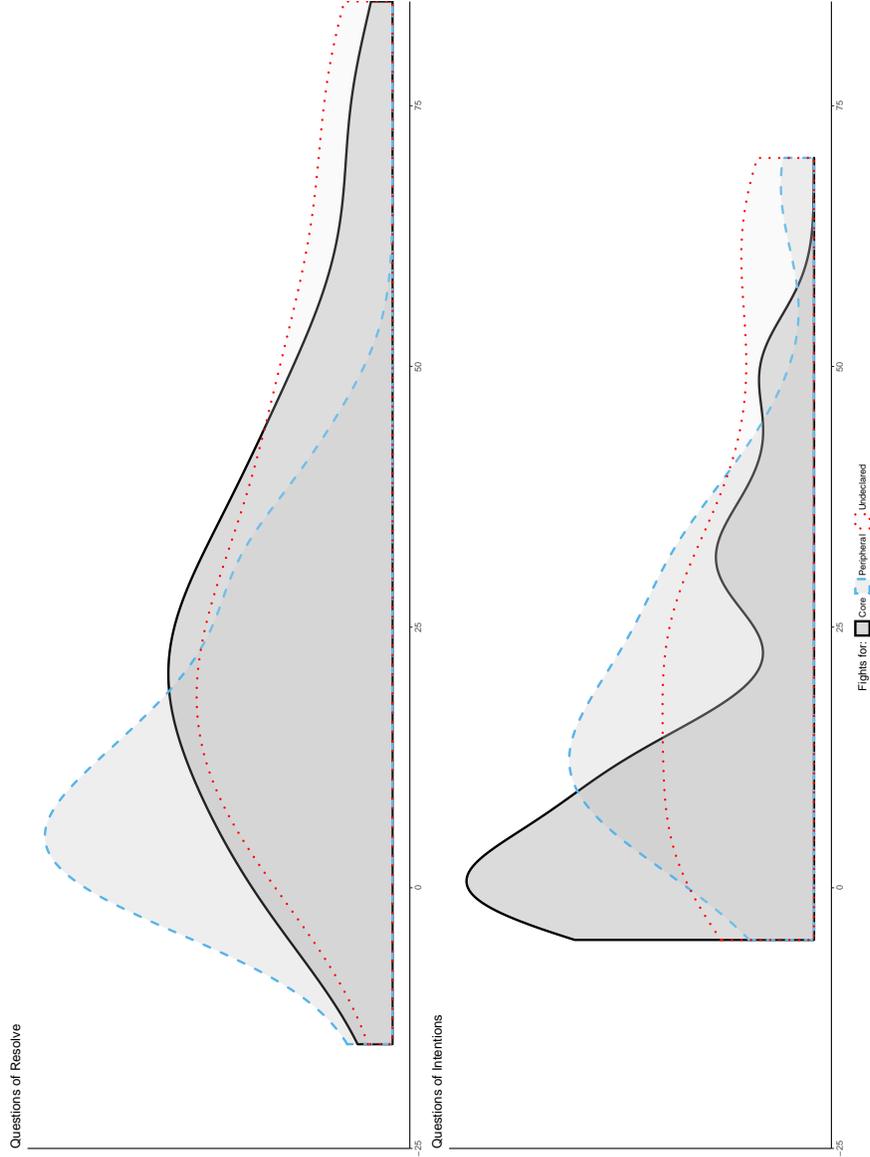
Panel B plots the distribution of responses. These plots increase our confidence for two reasons. First, they demonstrate that the group effects are not driven by a few outliers or polarized responses. Rather, when I expect that subjects will not update their beliefs, responses are amassed around 0, demonstrating that most subjects did not alter their beliefs. By contrast, when I expect subjects to update their beliefs, responses are distributed across positive numbers.

Figure 11: Changing Responses Across Time

(a) 95% Confidence Intervals.



(b) Distribution of Responses.



Responses are measured in changes in % pr. True from the second to third phase of the experiment. They are broken down into 3 treatment groups. Distribution of responses plotted above. Mean with 95% confidence plotted below.

## E Appendix C: Text Responses

We present responses to the question:

If other countries have aggressive intentions, they have strong incentives to hide them. So why should we believe anything that their leaders say? In your work, do you consider what foreign leaders say when you evaluate their interests? If so why?

categorized by the theory they support.

### Consistency Between Words and Deeds

Trust is at the core of all diplomatic efforts. Analysis of known data will indicate how much trust can be afforded to the rising power based on historical and current perspectives and actions.

Yes, it can give indications of intentions, to at least some degree.

I think we need to listen to what we are being told with a critical mind. That does not mean we automatically assume someone is lying to us, nor does it mean we take them wholly at their word. The closer we can develop diplomatic and security ties with a country, the harder it is for them to lie to us, but the more obvious to the other country if we think they are lying. What state leaders say is useful to a country assessment, but must be balanced with other information sources and historical context.

It is not possible to be certain that the target country is being truthful. Precedent/history, leader personality, the country's political system and other variables all contribute to this uncertainty. Because the decisions involved include people who are by their nature at least partially irrational and influenced by circumstance, ego, threats etc, every situation is different and it is not possible to make a generic assessment; doing so risks making significant errors. / So we should not necessarily "believe anything they say" but we must consider everything they say.

When assessing a situation in real life, there is more information than just the declarations of a country. Assessing the all intelligence available for a country would give a better basis on whether we should believe what they say. Regardless of whether what they say is truthful or misleading, it is important because a country would not say anything without first giving it careful thought. We, as analysts, must understand what that thought was so we can interpret what they say.

We should never simply “believe” anything said by foreign leader. We can never rule out that a rising power might have “aggressive” foreign policy objectives. That being said, if a nation’s leaders are dishonest about its intentions, then we won’t know that it has “aggressive” long term objectives until that nation is, in fact, aggressive. We cannot prevent a sovereign nation from building its military without evidence that it intends to use that military to violate international laws or undermine our own objectives (or that of our allies)

Yes, because I still think (as I did here) that it is possible to see undertones of defensive or offensive interests. Furthermore, I think it is important to see how their intentions when said line up or fail to line up with their clear international actions, as indicators of their sincerity.

Obviously actions are generally more indicative of underlying intentions (and in the application of international law are generally considered to be a stronger indication of a nation’s position than its statements). However, words can also be quite indicative; other than in full blown war scenarios, international disputes are fought out in multiple spheres, including diplomatic and even public relations ones. Therefore, it could be assumed that nations will attempt to claim a moral position and avoid lying or backflipping. It is for this reason that broad phrases such as acting for its ‘core security interests’ are used. This enables a moral stake to be taken (its reasonable for all countries to be concerned for and, to an extent, act to ensure security for its citizens and interests) and such a broad remit allows it to undertake a wide range of actions under that guise while seeking to argue that its objectives remain consistent. / / As such, statements that contain or allude to broad, subjective goals or principles can be taken to indicate that those making the statements may want to stake a moral claim yet preserve their ability to take a wide range of paths without diverting from the apparent broad moral principle.

There is seemingly always a gap between a given nation’s declaratory policy and their actual pursued policy, but that is not to suggest there is no value in the declaratory. Other sources of information need to be brought to bear (intelligence, domestic politics, past behavior, strategic culture, etc.) in order to often tease out basic strategic truths contained in declaratory policy.

Words do matter, but it really depends on the history with the nation and the setting. If there is a history of truthful statements and follow-through on measures, then the words of leaders are easier to believe. If the opposite is true, then words are much more difficult to believe.

Yes. There is always more information on which assessments can be made. Historic context and cultural awareness can help determine if statements are truthful or not. In any case, statements are useful in determining intent.

It is absolutely vital to believe what they say - unless and until facts on the ground contravene what was uttered; as well, one must invariably Trust, but Verify.... / / Yes it was. It gave a benchmark which then factored into a much larger tapestry of facts, informed

judgment, history, geopolitical imperatives, and strategic national interest, which = when synthesized - helps determine foreign policy.

Anything the target country says is useful, either through what is actually said, or from reading between the lines. Where a country knowingly spreads falsehoods it utilises diplomatic capital. As in this scenario once military action was taken trust in future actions was reduced leading to greater uncertainty and instability.

## **Neither Confirming Nor Dis-confirming**

It is usually what a country is not saying that a state must be most mindful. Hearing what another country is saying is important, but what is it leaving out of the equation? We must be sure to ask all of the right questions.

It is useful, but the cultural-historical-political-economic (esp. resources) context is often more telling.

Cheap talk is often useful; especially if you don't have much else to divine intent

## **Diplomacy is not Useful**

Depends upon the country - but honestly, what's said is rarely a useful basis for assessment.

We shouldn't believe anything they say. / / It is useful to a certain extent, but every nation has worked own self interests in mind.

## **Diplomacy is Useful Because of the Tone of the Meetings**

What the target country says is secondary to how the country says it. The message itself will be self-interested, biased and perhaps false. But how the message is promulgated and to whom the message is actually intended may reveal the true intentions of the message.