

Chapter 4: Elite Survey Experiment

July 12, 2017

1 Introduction

In this chapter we address the micro-foundations of my theory. We ask: do real world, individual policy-makers use the logic of consistency between words and deeds to analyze the intentions of rising powers?

It is difficult to answer this question with observational data. What rising powers say during diplomatic meetings, their choice of military targets, and who they choose to meet with are all strategic decisions. Power transitions last for decades and changing factors in military spending, regime type and other variables over time may drive different assessments as will changes in the institutional make-up of the declining power. The assessments that foreign policy professionals make may be confounded by select incentives that lead them not to report, or over report certain types of assessment. It is also difficult to know if the available de-classified information well captures assessments that were made.

To overcome all these issues, I present evidence from an elite survey experiment I administered to 93 real-world foreign policy, intelligence and defense professionals. The survey reflected an National Security Council (NSC) war game. I presented subjects information about a fictional rising power named Bandaria. Subjects were randomly assigned information about Bandaria's military behavior and diplomatic reassurances, then asked to assess Bandaria's long-term intentions as information was revealed.

Using random assignment on an elite sample I can address the main measurement and confounding issues found in observational data. I conclude that the elites in this sample, who faced a realistic scenario that reflects their basic work-functions, were less concerned about Bandaria's military behavior when it was consistent with diplomatic statements, *because it was consistent*.

As with any experiment that analyzes a convenience sample, I cannot extrapolate beyond the sample or the conditions of the experiment. The goal here is to show the internal validity of my theory. In later chapters I demonstrate that these basic findings are consistent with

broad trends in real-world assessment processes.

First, I describe the sample. Second, I present the vignette. Third I measure and test the hypotheses presented in my theory chapter. Fourth, I present supplementary tests designed to explore the validity of alternative theories. Finally, I draw inferences and present substantive implications.

2 Sample

Subjects were mid-level foreign policy elites from Australia and the United States. Subjects were eligible if they had briefed a Deputy Assistant Secretary, Congressperson or similarly ranked official on foreign policy issues. The solicitation email outlined these criteria. Subjects were asked sample inclusion questions at the end of the survey to ensure they met the elite sample frame. The survey was administered online through Qualtrics.

Subjects were solicited in two ways. 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.¹ Employers include the State Department, Office of the Secretary of Defense, various military and civilian intelligence agencies and staff for the Armed Services Congressional Committees.

Through these sampling processes, 139 subjects answered at least one question, 131 completed the survey, and 93 passed both attention checks and sample inclusion questions.

¹To be clear, I did not ask the people I knew directly to take the survey, only to distribute it through their network.

I analyze these 93 responses below. I describe the sample process and present summary statistics in Appendix C.

I wanted subjects to use their professional skills and experience to inform their answers. However, if they believed the scenario reflected a specific real-world case, they might have imported information from that case. Several steps were taken to ensure that subjects both had enough information to take the instrument seriously but responded to information in the instrument rather than real cases. First, the instrument was developed in collaboration with a retired Intelligence Director to make sure it included realistic information. Second, I administered the survey towards the end of the American presidential primary. The major interstate incident during this period was Brexit—an event unrelated to power transitions. Third, I chose a fictitious scenario with fake names and a map to avoid effects that arise when analysts think the scenario is a real-world case. 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. At the end of the survey I asked subjects if they thought the scenario reflected a real case. Most did not identify a case, and there was wide variation when they did.

2.1 Relationship Between Sample and Population of Interest

I address three concerns about inferences that I can draw about state-level assessments from my sample of mid-level foreign policy professionals. First, research shows that state inter-agency dynamics and select incentives influence decision-making ([Allison and Zelikow 1971](#)). Policies reflect power and veto points, not individuals' desires. These organizational concerns do not apply in my case because my dependent variable is beliefs, not policies. Beliefs form independent of interests. In the instrument, I ask only about the assessment of facts and not policy decisions based on those assessments.

Second, scholars claim psychological biases lead to analytical errors. These theories assume a rational baseline and explain how particular individuals (usually leaders) in the

decision-making process deviate from that rational baseline (Yarhi-Milo 2013; Mercer 1996). Unlike these theories, I am researching the rational baseline. Indeed, I find evidence for some biases that the literature theorizes. In my sample, these biases shift assessments in both directions. If I am right, then the average effect of consistency between diplomatic statements, military events and beliefs should be observed across the sample as a whole. These biases simply create error around my baseline expectation.²

Third, one might worry that mid-level policymakers may not reflect how cabinet members process information. 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. Thus, it represents a future pool of participants. 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 superiors see.

3 Vignette

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

Phase 1 records a baseline where all subjects receive the same prompt and 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 in preparation for that meeting. All are given a vignette that includes information about the trajectory of Bandaria’s military spending and economic growth, socio-political environment, trade and IGO networks and geopolitics.

²Since all I do is shift the rational baseline, I view theories of bias such as Yarhi-Milo (2013) and Mercer (1996), as complimentary with my account. Further work could analyze these accounts using my theory as the rational baseline.

The vignette includes a map of Bandaria and its surrounding countries depicted in Figure 1. 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 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.

The baseline allows me to analyze how different subjects bring different prior beliefs and assessment methods into the experiment.³ While there is much variation in initial assessments, a balance test confirms that all treatment groups begin with roughly equal priors.⁴

In *phase two*, subjects are randomly assigned into a cheap-talk *diplomatic treatment* where Bandaria justifies its military expansion to serve either: (1) security interests in the Lebang Bay, or (2) ethnic interests in Greywall in Arcadia. I also include (3) a control condition, where the diplomatic meeting calls for confidence building measures. I include two treatments to make sure diplomacy, and not the particular issue (security or ethnic based grievances), is doing the causal work. As an example, the text of the ethnic treatment is:⁵

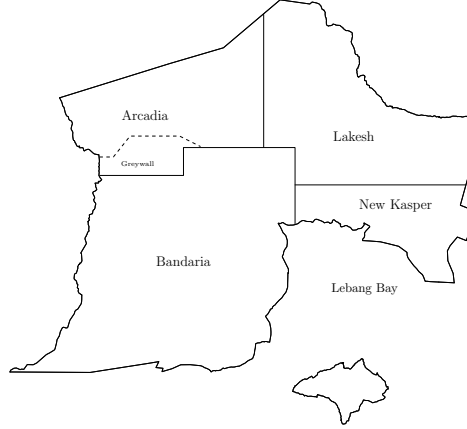
In a private meeting, the American president asked the prime minister of Bandaria to explain Bandaria’s military spending. The Prime Minister replied: “Grave injustices have been done to ethnic Bandarians. We have a long history of supporting our Bandarian brothers in Arcadia. Ethnic-national concerns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

³Summary statistics are included in Appendix C.

⁴A second balance test confirms treatment groups are well-balanced on biographical features.

⁵As the text in Appendix B shows, the different treatments are similar. There are only 10 different words between them.

Figure 1: Bandaria and its Neighbors



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.

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 1. Letters represent (s)ecurity

and (e)thnic treatments, and a (c)ontrol. 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 rhetoric and action. To that end, I pool consistent ($eE+sS$) and inconsistent ($eS+sE$) groups in the main analysis. I separate these groups in robustness checks. Further, in some specifications I test the difference within subjects across time. In these cases I am interested in subjects after they receive the diplomatic treatment in phase 2, and depict these groups s , e , c .⁶

Focusing on conflict provides the toughest test of my theory. Existing rationalist scholarship argues that military spending and interventions are a sure sign of aggressive intentions and should raise alarm (Mearsheimer 2001). If I am right, then policy-makers will not update their beliefs about rising powers following violent military actions that are consistent with diplomatic statements. Existing psychological explanations argue that analysts update based on impressions formed during meetings (Yarhi-Milo 2013; Holmes 2013). However, they would not predict that policy-makers who observe the exact same meeting update their beliefs differently depending on whether or not the information presented is consistent with military behavior.⁷

Table 1: Treatment Groups

	Ethnic Military Dispute	Security Military Dispute
Ethnic Meeting	eE : Consistent Ethnic	eS : Inconsistent
Security Meeting	sE : Inconsistent	sS : Consistent Security
Control	cS : Control	cE : Control

⁶After manipulation and selection checks, consistent = 29, inconsistent = 31, control = 33.

⁷Alternatively, these theories may say that a transcript of a meeting would have no effect. This still implies no difference between consistent and inconsistent groups.

3.1 Attention Checks

Every survey page contained summaries that updated as new information become available. Subjects could also access all of the information for review via a PDF link. To help subjects process information as they read, they were asked to write long-form assessments of Bandaria’s long-term intentions at different stages. The median character length for the first text response was 509 characters.

After the survey, subjects are asked to recall which country was invaded and how the Bandarian prime minister justified military spending during the diplomatic meeting. 73% of respondents correctly answered these attention checks. The remaining 17% were omitted from the sample. Subjects were also timed. Compared to an Mturk pilot sample, the elites took 50% longer to complete the survey (29 minutes) on average and twice as many passed the attention check. This gives me strong confidence that the elite sample paid attention.

3.2 Confounding Effects

Here I highlight how I dealt with two potential confounding effects in the design. Studies have shown that subjects are more likely to provide answers they believe to be desired by the researcher. A cursory look at existing research suggests that conflict generally should trigger desirability bias. Thus, this should attenuate the differential effect between consistent and inconsistent responses and work against my predictions. Nevertheless, I made sure to provide a vast amount of information before the subjects saw the treatment and did not discuss the importance of diplomacy in the prompt or baseline vignette.

Others may be concerned that my vignettes generates information for subjects other than my intended treatment (consistency). For example, if a claim of nationalism, or security is related to underlying factors (such as democracy), then the treatment may induce feeling about these factors that drive the results. I account for this bias in two ways. First, I provide information in the baseline vignette to hold constant many possible confounding factors.

Second, I allow diplomatic messages to include security and ethnic nationalist motives and military interventions to include these motives as well. If confounding was an issue, it should differentially effect these two types of treatments. We see that it does not.

4 Response and Measurement

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, the DIA and Hill staffers frequent describe probabilities but have no official standard. I did not want to favor one group over another. Second, there is much debate about what each category means. 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 (75%, 25%). But I wanted to avoid trivially small choices that would distract subjects. Using incremental measures and reducing the number of labels addresses potential biases that feeling thermometers may introduce.⁸

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 on the responses of a non-random sample conditional on the responses observed and the independence of the in-sample randomization.⁹

⁸See discussion between [Broockman, Kalla, and Aronow \(2015\)](#) and [Lupton and Jacoby \(2016\)](#) for problems and resolutions to feeling thermometers.

⁹The p-values are interpreted much like those found in t-tests. For an explanation of why they are superior to t-tests see [Ludbrook and Dudley \(1998\)](#). Using a t-test instead of the permutation test only

To test hypothesis 1, subjects are asked:

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

I estimate differences in mean responses to **A** between the treatment groups and the control condition after phase 2:

$$e \cup s > c \tag{1}$$

where the union implies groups are pooled in the estimation.

I further expect that subjects should coordinate their beliefs about what the rising power's limited intentions are based on the content of diplomatic messages. I present subjects with a click-map of Bandaria's region and ask:

B: In the last question you were asked to think about a most likely target. Click on the map where that most likely target is.

I expect that the click locations should correspond with the treatments receive for groups s , e . However, the control group (c) should be dispersed across the map.

Hypothesis 2 tests if inconsistency leads to alarm about long-term intentions. To test it, I ask:

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

Hypothesis 3 tests if inconsistency leads subjects to mistrust the rising power's diplomacy. To test it, I ask:

D: We can trust what the Bandarian Prime Minister said about Bandaria's long-term intentions.

improves my interpretation of any result reported in the paper.

I analyze questions **C** and **D** in two ways. I predict that subjects who observe inconsistent behavior will be more alarmed than those that did not. Thus, I estimate differences in group means across the consistent and inconsistent groups at phase 3 of the experiment:

$$eS \cup sE > eE \cup sS. \quad (2)$$

I also estimate differences in responses across phases 2 and 3 of the experiment holding individual beliefs constant. I predict subjects that observe inconsistent behavior should update their beliefs more than those that receive consistent behavior:

$$(eS \cup sE)_3 - (eS \cup sE)_2 > (sS \cup eE)_3 - (sS \cup eE)_2. \quad (3)$$

The subscripts (2 and 3) reflect responses after difference phases of the experiment. Although I cannot test a null hypothesis, I also expect that subjects who observe consistency do not update a great deal. Formally: $(sS \cup eE)_3 - (sS \cup eE)_2 \approx 0$.¹⁰

5 Results

Figure 2 presents results for hypothesis 1. The top of Figure 2 plots click-map responses. Panel (a) plots pre-treatment responses after phase 1. Panel (b) plots responses after the diplomatic treatment in phase 2. Pink circles received ethnic treatments, blue crosses received security treatments and black boxes received controls. As I expect, before treatment subjects from all treatment groups are well dispersed across the map. By contrast, after treatment responses reflect the content of their diplomatic justifications in almost all cases. Notably, the control group remains well dispersed before and after treatment.

The bottom panels in Figure 2 plot responses to question **A**. The lighter mass received a diplomatic treatment that contained a reassurance (group mean is the solid line). The darker

¹⁰I present alternative comparisons to all these tests between consistent message and the control groups in Appendix C. The results are substantively same.

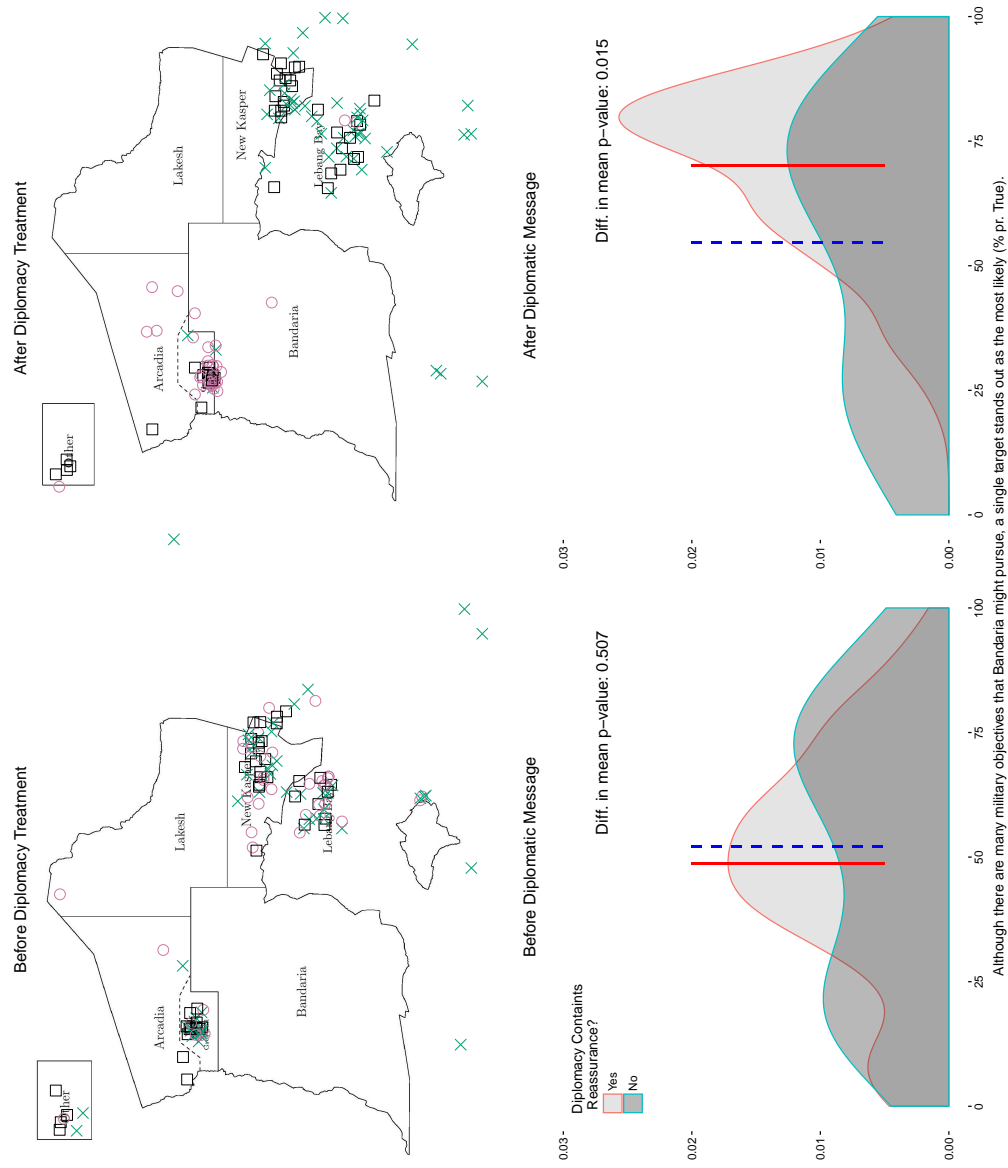


Figure 2: Does Cheap Diplomacy Influence Beliefs About Limited Aims?

mass received the control (group mean is the dashed line). Panels (a) plots responses before subjects received their diplomatic treatment, panel and (b) plots results after the diplomatic treatment. Before treatment the group means are the same.¹¹ After treatment the group means are different. Those that received treatment (rather than the control) became more confident that they understood what Bandaria wanted. A permutation test confirms that the means of treated and controlled subjects are different post-treatment with 98.5% confidence.

Hypothesis 2 examines how inconsistent behavior triggers concern about long-term intentions. Figure 3 plots the distribution of responses disaggregated by treatment. The darker mass observed a consistent diplomatic message and military intervention (group mean is dashed line). The lighter mass observed inconsistent treatments (group mean is solid line). Panel (a) presents pre-treatment results and panel (b) presents post-treatment results. Before subjects observed military interventions, the mean of both groups is not statistically different. The average respondent was more optimistic than not about Bandaria's long-term intentions, but the responses are close to complete uncertainty (50%). After the military intervention treatment, subjects that observed inconsistent behavior were concerned about Bandaria's long-term intentions. Subjects that observed consistent behavior remained uncertain. A permutation test confirms there is no difference in treatment group means pre-treatment but a significant difference post-treatment with 99% confidence. I infer that subjects who received inconsistent treatments grew worried about Bandaria's long-term intentions compared to those that observed Bandaria fight for what it said that it wanted.

Hypothesis 3 tests how consistent behavior affects trust. I plot the density of responses to question **D** before and after the military intervention treatment in Figure 4 disaggregated by consistent and inconsistent treatment groups. Before the military intervention treatment, the group means are nearly identical. Afterwards, subjects that observed inconsistent behavior are deeply mistrustful. Those that observed consistent words and deeds did not, on average, update their assessment. A permutation test confirms the means of these groups

¹¹Although the distributions are shaped differently the means are the same.

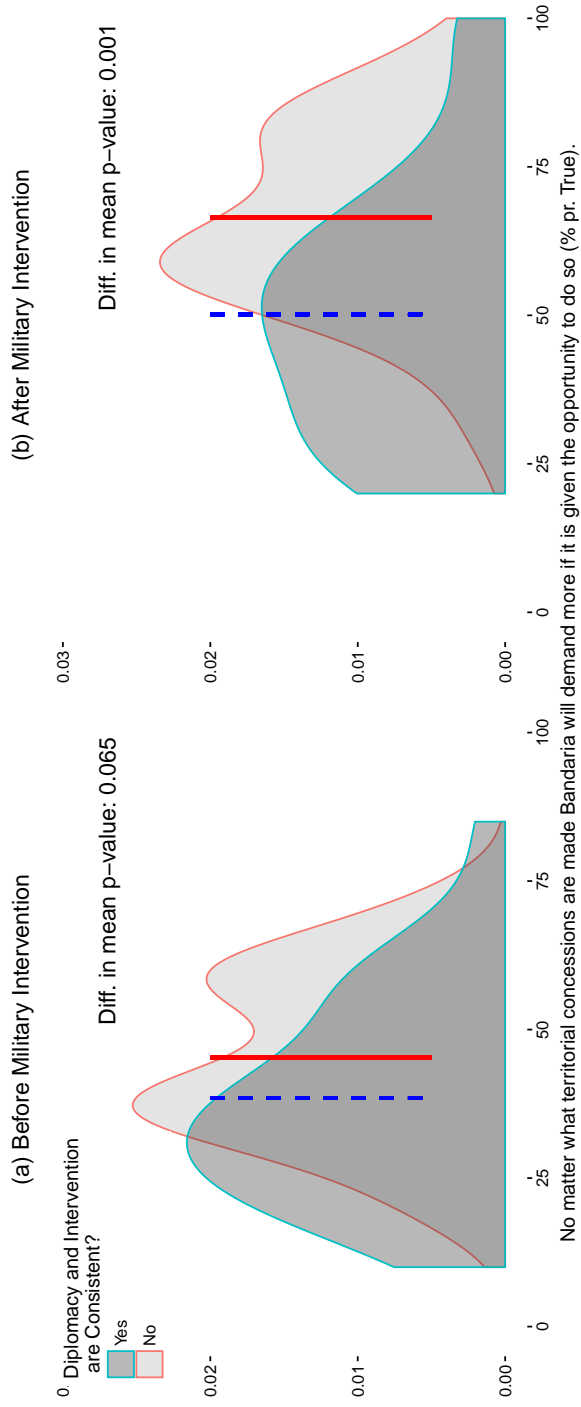


Figure 3: Does Inconsistent Behavior Alarm Subjects?

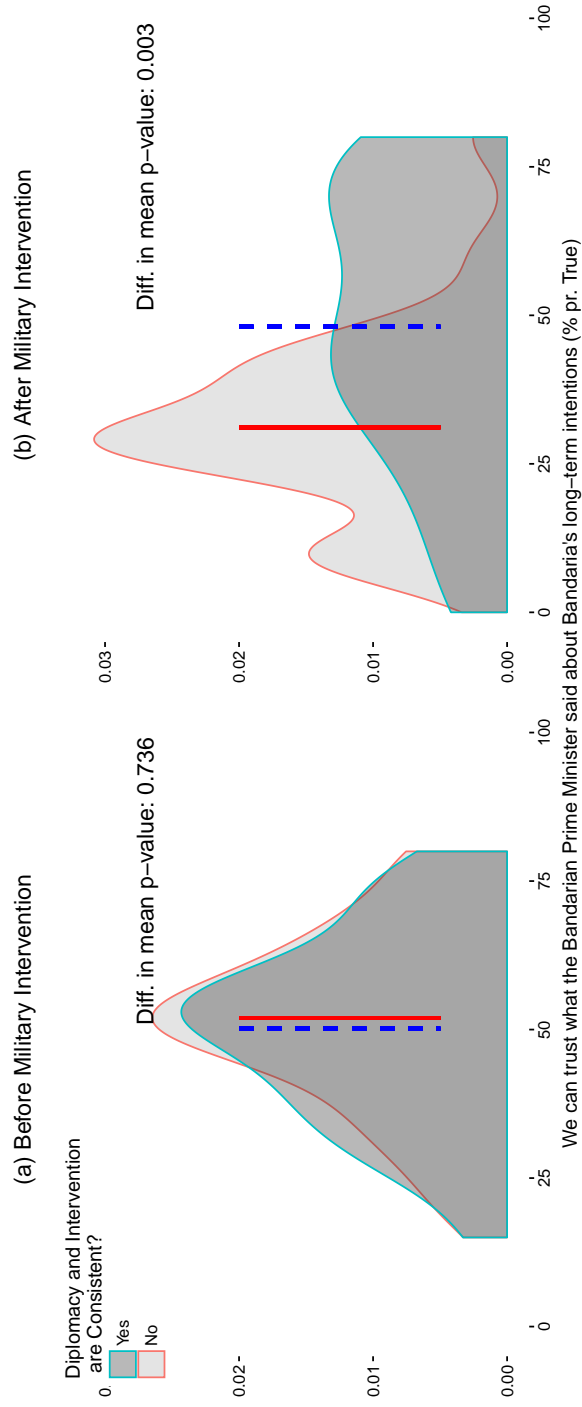


Figure 4: Does Inconsistent Behavior Lead to Less Trust?

is different with 99% confidence post-treatment. I infer that subjects who received inconsistent treatments grew mistrustful of Bandaria compared to those that received consistent treatments.

To further demonstrate the treatment effect, I analyzed how individual subjects adjusts their assessments across time in Figure 5. Panel (a) plots results to questions **C** (hypothesis 2). Panel (b) plots the results from question **D** (hypothesis 3). The x-axis plots the change in respondents' answer to the same question between phase 2 and 3. Responses can range from -100 to 100. The darker mass received consistent treatments and the lighter mass received inconsistent treatments.

As I expect, the consistent group is amassed around 0 in both cases. A permutation test cannot rule out that the consistent group mean is statistically different from 0 in either test. By contrast, the inconsistent group is dispersed broadly across positive ranges with a mean above 40 in panel (a), and negative ranges for panel (b) with a mean below negative 25. A permutation test confirms that the way the consistent group updated was different from the inconsistent group in both cases. I infer from these results that subjects did not alter their beliefs when they observed consistent violent behavior. However, subjects that observed inconsistent behavior become more concerned about Bandaria's long-term intentions and were less likely to trust what the Bandarian prime minister said in future diplomatic meetings.

5.1 Robustness Checks

In Appendix C, I present OLS regressions where the dependent variable are the responses to questions **C** and **D**. The right hand side variables include the treatment (consistency) and one of the following controls: sampling method, diplomatic message that was received, duration of the experiment, subject's work function, employment sector, military service record, nationality and seniority (measured by the most senior person they have briefed). In every model, the treatment is significant and consistent with the above results. However,

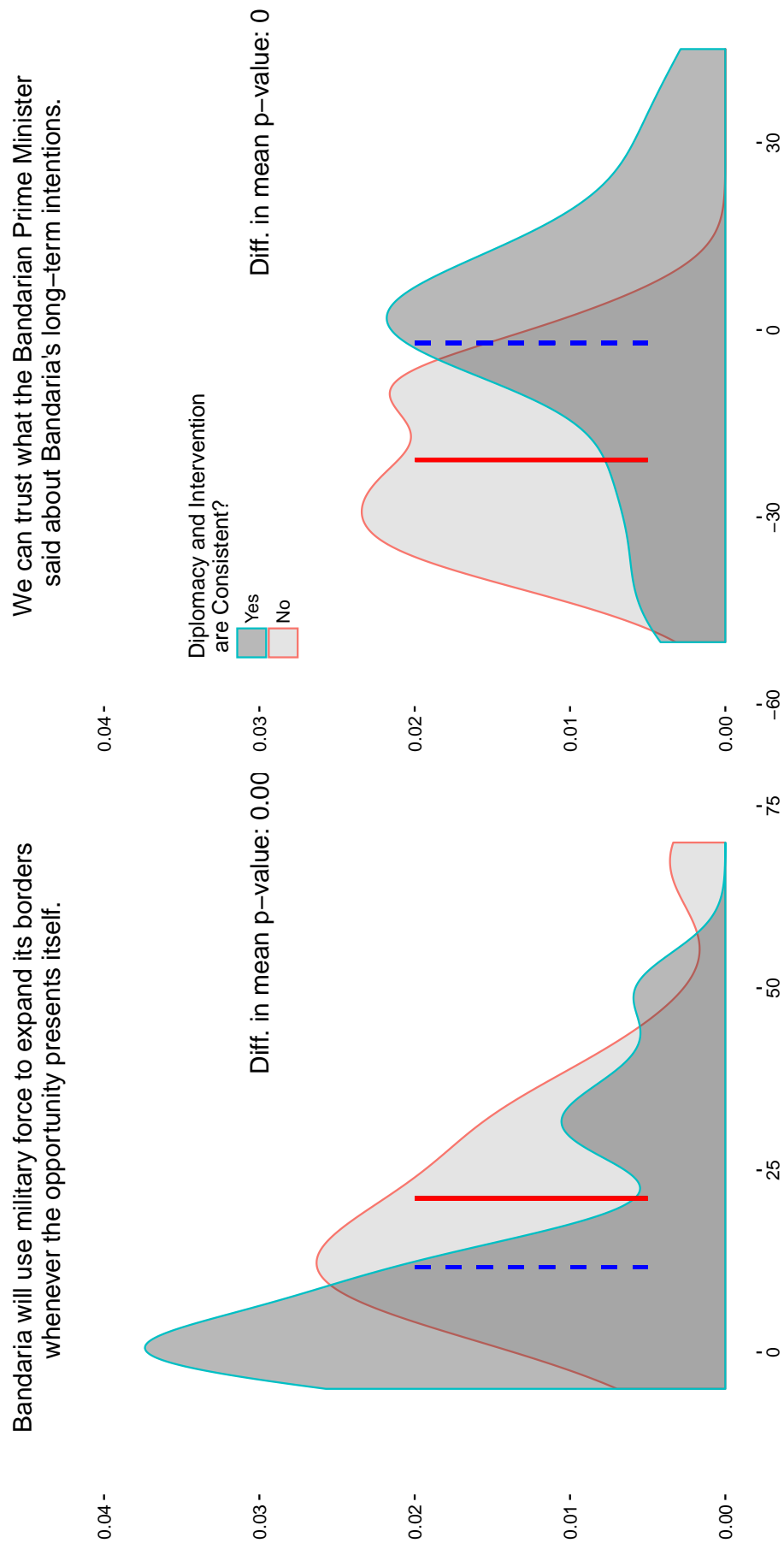


Figure 5: Does Inconsistent Behavior Lead Subjects to Change Beliefs Across Time?

the controls are not significant and do not confound the treatment effect. An analysis of the covariance suggest that the treatment does not covary with the controls.

Finally, my theory predicts that elites do not suffer from specific biases that differentiate them from the public. A pilot survey of Mechanical Turk workers who received college education yields similar results.

5.2 Further Analysis

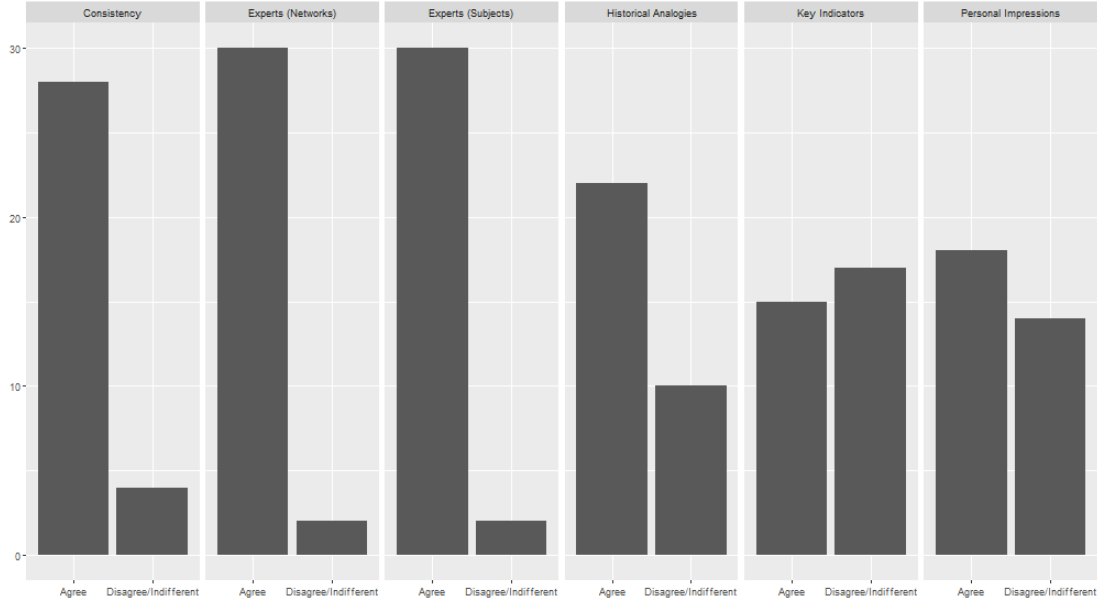
The hypotheses I tested here are not well explained by other theories in international relations. The differential effects of consistent diplomacy, as opposed to no diplomacy, and inconsistent diplomacy provide strong support for my theoretical mechanism. Of course, the world is complex. Several mechanisms may be at work simultaneously. To shed light on these alternative theories, I embedded post-survey questionnaires into the design and some additional questions in the main experiment to test alternative theories. Although the results are drawn from a small convenience sample, they shed light on how real-world decision-makers approach problems.

A variety of theories explore how analysts frame their assessments and approach analytical problems. Prominent international relations theories grounded in psychology and sociology suggest that policy-makers rely on historical analogies, and personal impressions to form their assessments. Organizational theories (both rational and sociological) suggest that policy-makers rely on subject matter experts to form their opinions. Some in decision-sciences expect policy-makers to rely on key indicators (although this has not been directly extended to intelligence agencies). To see how analysts approach analytic challenges I included a post-survey questionnaire:

Analysts can approach problems in a range of ways. If you were asked to analyze the intentions of a target country that you did not know much about, a good way to approach the problem is to:

- Compare the target to historical cases that were similar and extrapolate from those cases.

Figure 6: How Policy-Makers Approach Problems:



- Read what the targets leaders have said about their intentions and determine if their behavior is consistent with what they say.
- Use a standard list of key indicators that suggest a state is aggressive. Then see if the target matched those key indicators.
- Seek meetings with counterparts in the targets government and form impressions of them.
- Ask country experts with extensive historical knowledge of the target and use their analysis.
- Ask country experts with extensive networks in the target country and use their analysis.

41 subjects from the main survey were randomly assigned to this questionnaire, 34 responded. Subjects were asked to select from a five point scale. Figure 6 summarizes the responses grouped by Agree (Strongly Agree, Agree) and Disagree/Indifferent (Strongly Disagree, Disagree, Neither agree nor disagree). The results clearly show two types of methods are preferred by subjects: consistency between words and deeds, and a reliance on experts. Historical analogies are also more useful than not. Personal impressions formed during diplomatic engagements and key indicators were not thought to be useful by about half of the respondents.

The results also show that analysts positively report on multiple methods. Typically,

organizational and state-level explanations are pitted at odds with each other. It is common that scholars will suggest that their method is the dominant way that analysts will process information. These results suggest that individual policy-makers may use a variety of approaches to deal with problems. One reasonable interpretation of this data is that policy-makers gather expert opinions and then use them to evaluate consistency between words and deeds.

A second group of theories makes predictions about the effectiveness of different types of communication. Audience cost theory predicts that public statements are more important than private diplomacy. Theories of cognitive biases predict that diplomacy is important because it leaves specific personal impressions with policy-makers. We test these theories with a second post-survey questionnaire that was randomly assigned to 40 participants.

Q1: Think of times when you evaluated the intentions of a foreign counterpart. In these cases, would you find a private meeting with government officials from the target state to be useful for your assessment?

Q2: When you are unable to meet with foreign counterparts, are any of the following sources useful to you to supplement the missing information you would have gathered during these meetings:

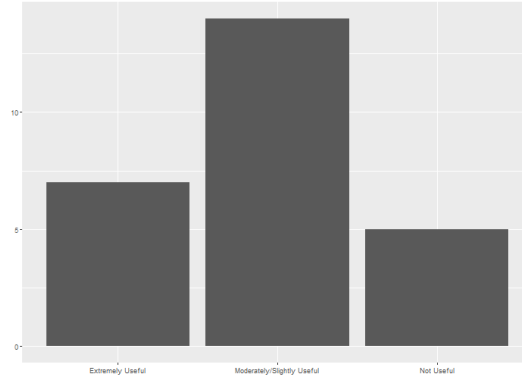
- A written transcript from a private meeting between the target's leader and a senior US diplomat where they discuss the target's long-term intentions in detail.
- A brief from a US diplomat that meets extensively with the leader of the target state.
- Speeches the target leader makes to their domestic public about their foreign policy goals.

Q3: Compared to other people in your profession are you a good judge of character?

Subjects responded to the first two questions with a seven point Likert scale from Extremely Useless to Extremely Useful. Responses to **Q3** were recorded on a 7-point scale from One of the very best, to one of the very worst.

Audience cost scholars argue that public statements are more credible than private statements because domestic and international audiences punish leaders that walk-back from public statements. I argue that private diplomacy is a vital part of the assessment process.

Figure 7: Are Meetings Useful?



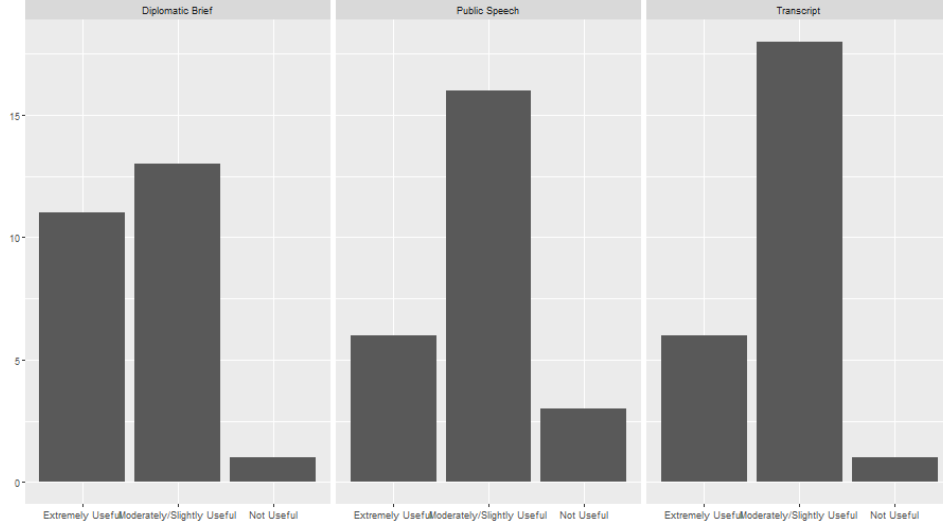
Q1 plotted in Figure 7, tests these competing conjectures directly. For ease of interpretation I combine the four useless categories together (Extremely useless, Moderately useless, Slightly useless, Neither useful nor useless), the two moderately useful categories (Moderately Useful/Slightly Useful), and present Extremely Useful separately. Very few subjects believed meetings were not useful. More thought they were extremely useful than not useful. Clearly, these policy-makers found private diplomatic meetings are useful to policy-makers.

The question, then, is why? My theory provides a rational explanation. However, theories of cognitive biases claim that leaders mistakenly rely on personal interactions to form their beliefs about the intentions of their rivals.

Although these theories differ in form. At the most basic level, they all expect that being present in meetings is a critical feature. Some go as far as to say that leaders make inferences from looking others in the eye. It would be surprising for these scholars to learn that analysts could replace a personal meeting with a simple transcript, or that arrogant leaders would trust the opinions of others in these meetings.

Q2, plotted in Figure 8, tests this basic proposition explicitly. All but one policy-maker in the sample believed that a written transcript was a useful substitute for an in-person meeting. One third thought a transcript was extremely useful. Subjects preferred a brief from a senior diplomat who regularly attended such meetings. Both of these results are surprising if the *personal* interaction during meetings generated information. The most reasonable

Figure 8: What Substitutes for Personal Interactions in Meetings?



interpretation is that policy-makers infer information from the content of what is said during diplomatic meetings, rather than impressions that they form of foreign counterparts.

Interestingly, public speeches are the worst substitute for private diplomatic meetings (although still quite good). One interpretation of this result is that private meetings convey different information from public speeches.

In a third post-survey questionnaire we posed the problem of cheap-talk to subjects directly:

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?

The question wording was explicitly designed to prompt respondents to say no. Thus, it presents a tough test of my theory. Only 18 subjects replied from subjects assigned to this group.¹² All the answers appear in Appendix C. 12 of the 18 provide clear support for my theory with 3 additional responses are not detailed enough to make a judgment, 1 supports theories of cognitive bias and 2 report that diplomacy does not matter.

¹²We reason that the lower response rate was due to long-form text responses rather than than click responses.

5.2.1 Discussion of Supplementary Results

A variety of supplementary test, provide qualitative support for my argument. Policy-makers have a clear preference for consistency between words and deeds over personal impressions and historical analogies. Policy-makers are conscience that diplomacy helps them make decisions. Even when posed with the problem of cheap-talk directly, they still say that diplomacy matters and explain that it is best used to verify events that unfold. Furthermore, policy-makers do not believe that in person meetings are necessary. Rather they can get as much information from expert briefs with people that attended meetings or even a transcript of the meeting itself.

There are many inferential challenges with these supplementary results. Yet it is surprising that over a variety of different questions and long-form responses designed specifically to push respondents to dis-confirm my theory, the responses are supportive. Further, I find no support for alternative theories from audience cost and cognitive bias theories in a variety of simple tests.

6 Inferences and Substantive Implications

The results suggest that policy-makers do not always respond the same way when a rival uses force. This experiment suggest that the context provided by diplomatic interactions is critical to how policy-makers interpret these events. When violent military behavior is consistent with long-standing diplomatic claims, analysts do not adjust their beliefs. By contrast if either diplomacy is absent, or diplomacy is inconsistent with what was said, then subjects will grow alarmed.

The magnitude of the effect is remarkable given subjects participated in an online study that took about thirty minutes with complex information and three batteries of responses. Indeed, it would not be difficult for subjects to miss critical information, or ignore the nuanced features of this experiment. These factors should have attenuated my main finding.

The fact I still find such strong results is a good sign that the effect is genuine.

The results are robust to several alternative specifications and robustness checks. Furthermore, I used post-survey questionnaires to rule out alternative mechanisms that are inconsistent with my main results but possibly driving an underlying effect.

To close, I consider the substantive implications of the results. I consider the hypothetical judgment that the mean subject from each treatment group would write about Bandaria's long-term intentions in the wake of a Bandarian invasion against a third-party. The briefs translate the percentage results from the experiment to the confidence scale used by the CIA in their assessments.

In the consistent case, the brief would read:

We are no more concerned about Bandaria's long-term intentions as a result of this military intervention. We still lack sufficient information to evaluate Bandaria's long-term objectives or its willingness to use force opportunistically. We will continue to observe Bandaria's behavior and engage Bandarian leaders about their intentions.

In the inconsistent case, the brief would read:

We are now more concerned about Bandaria's long-term intentions as a result of this military intervention. We now assess with moderate to high confidence that Bandaria's long-term intentions are aggressive and that Bandaria will use force opportunistically whenever the opportunity presents itself. We will continue to observe Bandaria's behavior but do not believe that future discussions with the Bandarian leadership will provide credible information about Bandarian long-term intentions.

Since some subjects brief NSC member including the Joint Chiefs of Staff, and the Secretary of State and Defense, these briefs reasonably capture the different summaries that the National Security Council would report to the president.

References

- Allison, Graham T. and Philip Zelikow, 1971. *Essence of decision: Explaining the Cuban Missile Crisis*.
- Broockman, David, Joshua Kalla, and Peter Aronow, 2015. Irregularities in LaCour (2014).
- Dafoe, Allan, Baobao Zhang, and Devin Caughey, 2016. Confounding in Survey Experiments.
- Holmes, Marcus, 2013. The Force of Face-to-Face Diplomacy: Mirror Neurons and the Problem of Intentions, *International Organization*, **67**(4), 829–861.
- Ludbrook, John and Hugh Dudley, 1998. Why Permutation Tests Are Superior to t and F Tests in Biomedical Research, *The American Statistician*, **52**(2), 127.
- Lupton, Robert and William G. Jacoby, 2016. The Reliability of the ANES Feeling Thermometers: An Optimistic Assessment.
- Mearsheimer, John J, 2001. *The Tragedy of Great Power Politics*, Norton, New York.
- Mercer, Johnathan, 1996. *Reputation and International Politics*, Cornell University Press.
- Strasser, H and C Weber, 1999. On the asymptotic theory of permutation statistics, Tech. rep., WU Vienna University of Economics and Business, Vienna.
- Yarhi-Milo, Keren, 2013. In the Eye of the Beholder: How Leaders and Intelligence Communities Assess the Intentions of Adversaries, *International Security*, **38**(1), 7–51.

A Appendix A: 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.

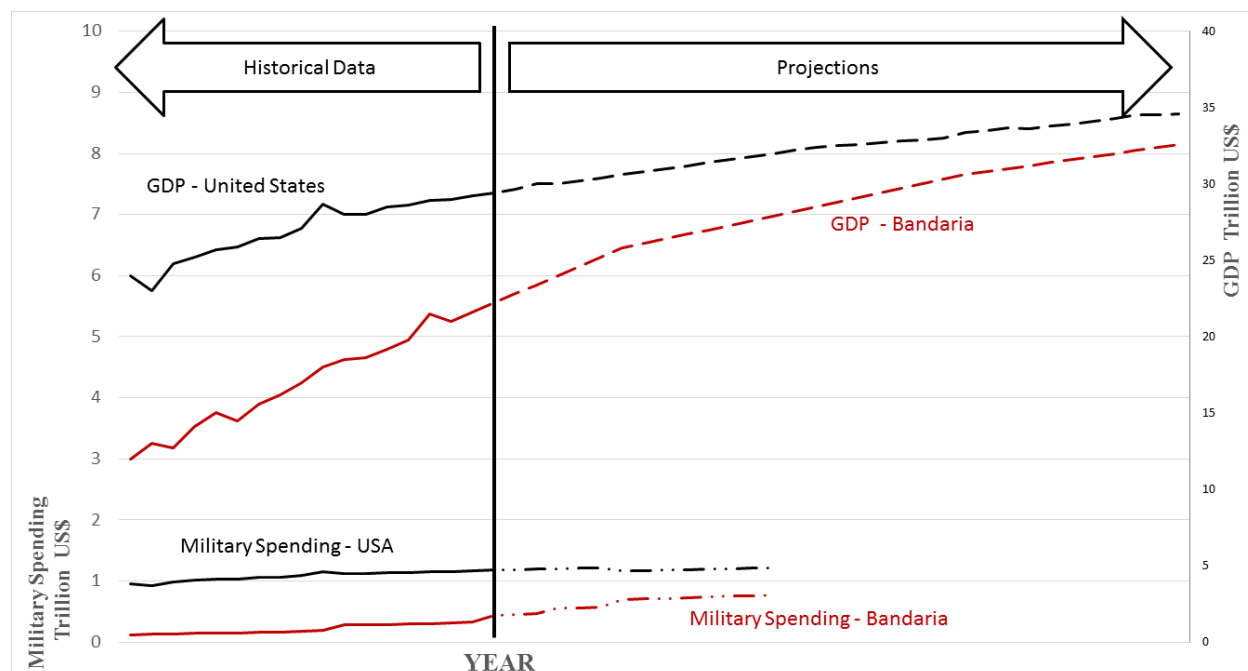
Evaluate the following information and provide an assessment of Bandaria's long-term intentions.

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 9: 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 10: 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

A.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

A.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.

A.0.3 Minutes of Diplomatic Meeting

A.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.

A.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.

A.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.

A.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

A.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.

A.0.9 Breaking news

A.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.

A.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.

A.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.

B Appendix B: 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.

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).

Solicitation:

The sampling method relied on a convenience sample of policy elites. As described in the manuscript, I used two sampling methods described. In the snowball sample, I asked my professional network to extend a solicitation email to subjects that they believed fit my sampling frame. The following text is the solicitation email sent to subjects from the snowball sample:¹³

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.

¹³The Solicitation for the institutions is similar, and therefore omitted.

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]

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,

Michael Joseph George Washington University

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 2 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 2: Dropout Rates Between Answering First Question and All Questions.

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) some 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.

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,¹⁴ and completed the survey, 32 failed at least one of the attention checks. Table 3 shows that they are well dispersed across treatment groups.

	Passed	Failed
Consistent	29	11
Control	33	8
Inconsistent	31	13

Table 3: Elites Who Passed Attention Checks?

Subjects in Reported Results

The results reported in the analysis describe the 93 subjects that passed both sample inclusion and attention checks.

Summary Statistics

The following tables breaks out treatment groups by various covariates.¹⁵ I broke out the tables to show that there is a good dispersion of covariates across the sample and most

¹⁴i.e subjects I am confident are elites.

¹⁵Note 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.

subjects responded to biographical information if they completed the survey.

Since I rely on in-sample randomization, it is difficult to extrapolate my findings to a broader population. However, the broad dispersion of the biographical information increases my confidence that I have captured a somewhat representative sample of the American foreign policy community. For example, 53% of the sample has some military service. Although this is higher than the American population, the military is over-represented in intelligence and foreign affairs communities. Furthermore, Table 5 shows that subjects with military service are mainly employed in Military Agencies (including both the Military and the Defense Intelligence Agency), but are also represented in civilian agencies, congress and the private sector analytical community.

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.¹⁶

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.

¹⁶Not all subjects answered all biographical information. The numbers therefore do not always add up to 93.

Table 4: Biographical Data By Treatment

(a) American Citizen:

	Consistent	Control	Inconsistent	Total
Yes	24	21	22	67
No	5	12	9	26
Total	29	33	31	93

(b) Military Service:

	Consistent	Control	Inconsistent	Total
Yes	14	20	15	49
No	15	12	15	42
Total	29	32	30	91

(c) Employment Sector:

	Consistent	Control	Inconsistent	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

(d) Primary Work Function:

	Consistent	Control	Inconsistent	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

(e) Highest Ranked Person You Briefed:

	Consistent	Control	Inconsistent	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 5: 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

Table 6: Attention Data By Treatment

(a) Duration of Survey:

Treatment Group	Median Minutes
Consistent	25
Control	29
Inconsistent	36
Full Sample	33

(b) Character Response:

Treatment Group	Median Character Length
Consistent	457
Control	498
Inconsistent	623
Full Sample	509

Regression Analysis

To further demonstrate the proper application of randomization, I report regression results that include the covariates above. 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:

$$Trust_i = \beta_0 + \beta_1 Consistent + \beta_k Control_j + \epsilon \quad (4)$$

$$Opportunistic_i = \beta_0 + \beta_1 Consistent + \beta_k Control_j + \epsilon \quad (5)$$

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. Sample Method: (1) Institutional Sample, (0) Snowball Sample
2. Message Type: (1) Security Message, (2) Ethnic Message
3. Work Function
4. Employment Sector
5. Military Service
6. Citizenship: (1) American, (0) Australian/New Zealand
7. Highest Ranking Official: Ordinal variable
8. 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 every single model. Further none of the controls are significant.

Table 7: Sampling Method

	<i>Dependent variable:</i>	
	Trust (1)	Opportunistic Force (2)
Consistent Treatment	16.760** (5.494)	-17.081** (5.133)
Institutional Sample	0.872 (5.737)	-2.663 (5.359)
Received Security Message	-3.820 (5.608)	-3.547 (5.239)
Constant	33.048** (5.421)	69.914** (5.065)
Observations	60	60
Adjusted R ²	0.111	0.124
<i>Note:</i>		*p<0.05; **p<0.01.

Table 8: Work Function

	<i>Dependent variable:</i>	
	Trust (1)	Opportunistic Force (2)
Consistent Treatment	21.653** (6.351)	-19.171** (5.867)
Policy-Making	-7.825 (7.127)	-6.791 (6.584)
Programmatic Work	-0.036 (7.291)	-12.660 (6.735)
Diplomacy	-12.693 (9.985)	0.260 (9.223)
Constant	32.149** (5.300)	73.216** (4.896)
Observations	59	59
Adjusted R ²	0.129	0.162
<i>Note:</i> *p<0.05; **p<0.01.		
<i>Note:</i> Baseline Category is Research/Analysis		

Table 9: Employment Sector

	<i>Dependent variable:</i>	
	Trust (1)	Opportunistic Force (2)
Consistent Treatment	17.827** (5.634)	-17.186** (5.262)
Military Agency	-3.475 (6.142)	-2.130 (5.738)
Political Party	-5.870 (12.913)	-10.127 (12.061)
Private Sector	-5.399 (9.570)	4.508 (8.939)
Other	17.681 (21.807)	-18.251 (20.370)
Constant	32.319** (5.141)	68.251** (4.802)
Observations	59	59
Adjusted R ²	0.100	0.114
<i>Note:</i> *p<0.05; **p<0.01. <i>Note:</i> Baseline Category is Civilian Gov. Agency		

Table 10: Military Service

	<i>Dependent variable:</i>	
	Trust (1)	Opportunistic Force (2)
Consistent Treatment	17.357** (5.408)	-16.690** (5.104)
Military Service	4.643 (5.408)	1.690 (5.104)
Constant	28.345** (4.656)	65.988** (4.395)
Observations	59	59
Adjusted R ²	0.136	0.131
<i>Note:</i> *p<0.05; **p<0.01.		

Table 11: American Citizenship

	<i>Dependent variable:</i>	
	Trust (1)	Opportunistic Force (2)
Consistent Treatment	16.631** (5.416)	-16.267** (5.088)
American Citizen	2.915 (6.399)	-0.102 (6.012)
Constant	29.060** (5.876)	66.524** (5.520)
Observations	60	60
Adjusted R ²	0.122	0.125
<i>Note:</i>		*p<0.05; **p<0.01.

Table 12: Highest Ranking Official

	<i>Dependent variable:</i>	
	Trust (1)	Opportunistic Force (2)
Consistent Treatment	16.468** (5.748)	-16.046** (5.418)
Rank of Official	-3.637 (3.947)	1.263 (3.720)
Constant	41.698** (12.578)	63.001** (11.856)
Observations	57	57
Adjusted R ²	0.136	0.123
<i>Note:</i>		*p<0.05; **p<0.01.

Table 13: Time to Complete Survey

	<i>Dependent variable:</i>	
	Trust (1)	Opportunistic Force (2)
Consistent Treatment	17.410** (5.292)	−16.765** (4.928)
Duration	4.569 (3.267)	−5.097 (3.042)
Constant	29.900** (3.777)	67.823** (3.516)
Observations	60	60
Adjusted R ²	0.148	0.166
<i>Note:</i>		*p<0.05; **p<0.01.

Alternative Counter-factual

In the manuscript, I compare the consistent group to the inconsistent group. I do so for two reasons. First, it matches the main interpretation of my hypotheses. Second, it ensures that subjects are not exposed to heterogeneous treatment effects because the same treatments in different combinations are provided to both groups.

An alternative interpretation of my hypotheses could lead a reader to expect that consistent group should be compared to the group that observed the control message, then a military intervention without diplomatic context at all. Below I re-run my analyses for H2, and H3 testing:

$$sS \cap eE > cS \cap cE. \quad (6)$$

The results are substantively the same as the results reported in the manuscript. One interesting difference is that the distribution of responses in the control group is significantly more dispersed than the inconsistent treatment group. To the extent that a wider variance in responses reflects greater uncertainty, this further supports my theory.

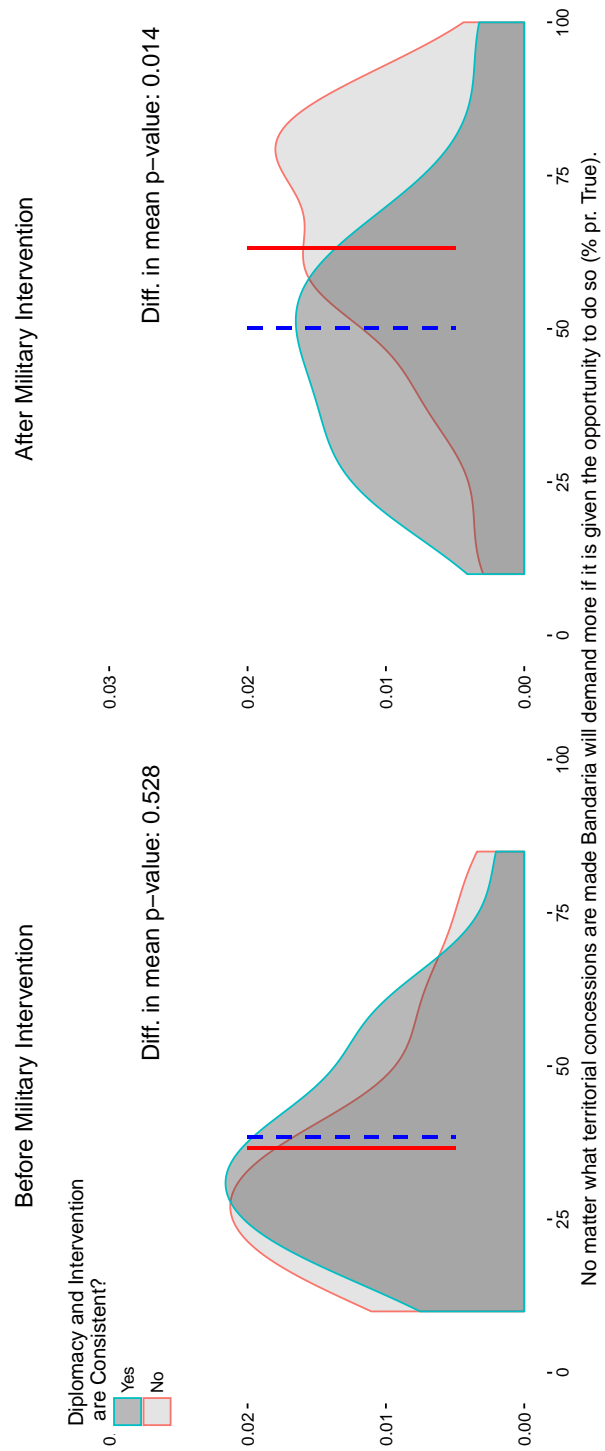


Figure 11: Does Cheap Diplomacy Influence Beliefs About Limited Aims?

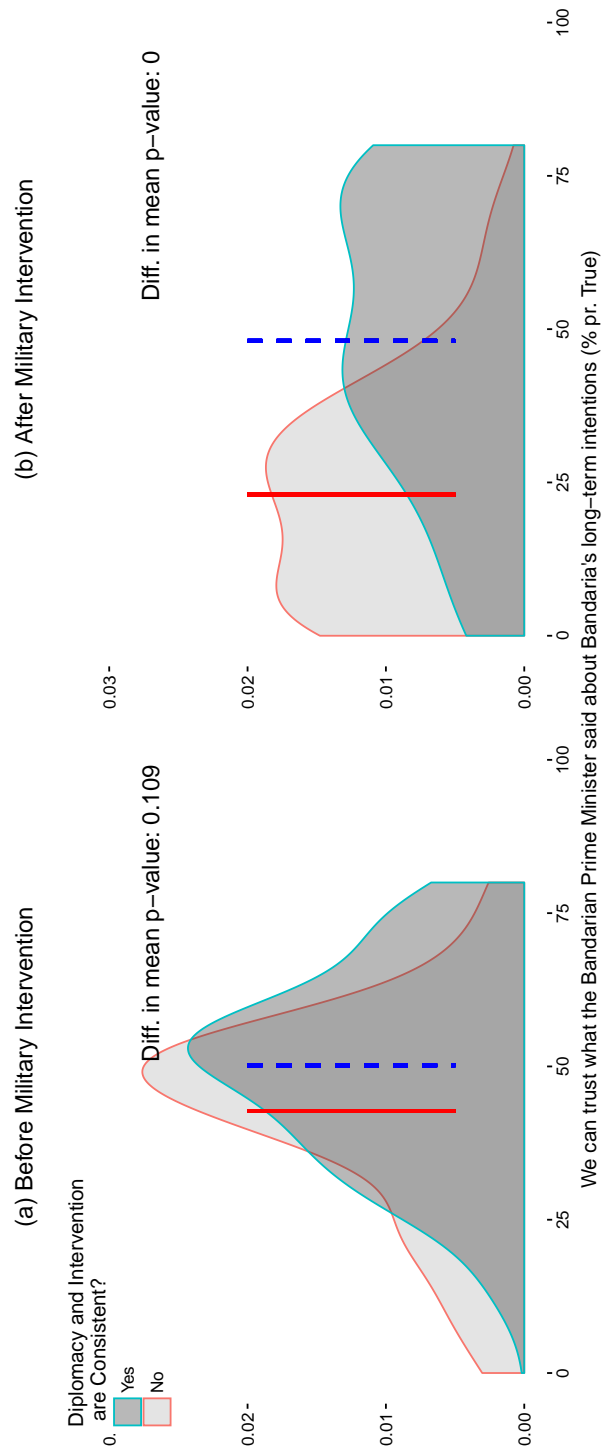


Figure 12: Does Inconsistent Behavior Lead to Less Trust?

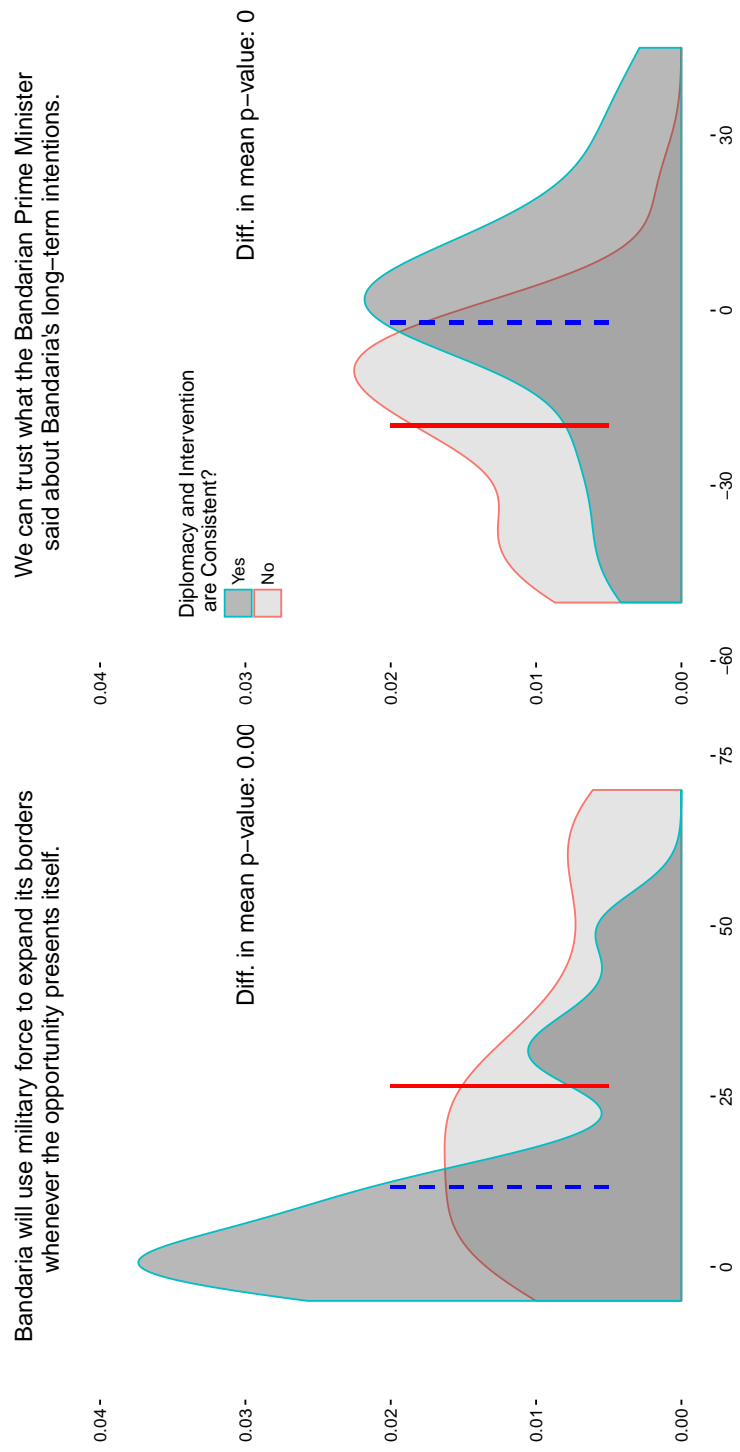


Figure 13: Does Inconsistent Behavior Lead Subjects to Change Beliefs Across Time?

C 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.