A motives-driven theory of power transitions

Michael F. Joseph
Yale University
mfjoseph@gwu.edu
Word Count: 14,784

September 14, 2019

Abstract

What explains patterns of competition and cooperation during power transitions? I develop a formal model to explain competition through changes in the declining power's beliefs about the rising power's intentions. In it, rising powers use costless diplomacy to credibly signal that their long-term intentions are limited by tying demands for specific concessions (e.g., Taiwan) to underlying principles (e.g., nationalism). I show that the rising power's costless diplomatic message at the beginning of the power transition persuades the declining power to delay preventive competition even if the rising power makes territorial demands. I also show these early costless promises help establish a delayed test of the rising power's costly demands that explains competition after a specific territorial demand that comes in the middle of the power transition. Declining powers only turn to competition once they discover that the rising power's original diplomatic message was dishonest. Thus, inconsistent behavior and not violent behavior alone triggers competition. I validate my theory using an elite survey experiment that simulates a National Security Council assessment of an emerging threat and randomly assigns diplomatic messages and military interventions to subjects. The subjects, real-world foreign policy professionals, trusted the emerging threat when military interventions matched diplomacy and mistrusted it otherwise.

NOTE TO READER: This is my job market paper. It summarizes the theory and causal evidence from my book project. For additional analyses and sample chapters from my book project please see the job market page on my website.

Introduction

What explains patterns of competition and cooperation during power transitions? Through history, an important factor has been changes in the declining power's beliefs about the scope of the rising power's long-term motives. For example, in 1932 British elites were unsure if Hitler was a "madman," bent on world domination or if Hitler's aims were limited. The British worried when they observed Hitler rapidly militarize, then annex the Rhineland, then parts of Austria and Czechoslovakia. But Hitler allayed their fears through diplomacy. During private exchanges, Hitler convinced three British prime ministers that he would peacefully integrate into European society once he satisfied German-nationalist goals. Only once Hitler violated the Munich Agreement did the British realize his aims were vast, and this realization triggered competition.

Seven years later this pattern repeated. Following a week-long diplomatic meeting with Stalin, Churchill noted that "poor Neville Chamberlain believed he could trust Hitler. He was wrong. But I don't think I'm wrong about Stalin."[1] With Hitler fresh in their minds, British and American leaders believed that Stalin's intentions were limited up until the Iran Crisis (1946). At that point the Western Allies realized that Stalin's intentions were expansive, and switched to containment. Even today, American policy-makers know that inferring "Chinese intentions is the single most difficult and important task we face,"[2] yet rely on high-level diplomatic meetings to make inferences about Chinese intentions; leading to Sino-American cooperation.

For nearly a century, three aspects of this pattern have puzzled rationalist scholars (Carr, 1964). First, researchers expect that the onset of power transitions carries the greatest risk of competition. At the beginning of power transitions, declining powers are most uncertain about the rising power's intentions because there has been no opportunity to cultivate a peaceful reputation (Kydd, 2005; Glaser, 2010). Further preventative war, and contain-

[1] Nicholson Diaries, 27 Feb 1945. Discussion about Soviet strategic intentions.
[2] Quote from interview with former CIA Director for Analysis, Mark Lowenthal.

ment are most attractive early on (Powell, 1999, 2006; Debs and Monteiro, 2014). As a result, if competition is the declining power's best strategy, it should come early or not at all (Schweller, 1992). *Why is cooperation and trust so common at the beginning of power transitions?*

Second, researchers do not know why cooperation eventually breaks down. Some link competition in the middle of power transitions to differential rates of shifting power or erratic disturbances in the balance of power (Kugler and Lemke, 1996; Krainin, 2017). But evidence suggests that declining powers realize that the rising power's long-term intentions are vast, and this change in beliefs triggers competition. For example, when Hitler violated the Munich agreement Chamberlain re-evaluated whether cooperation was possible. It was this change in beliefs about Hitler's intentions that triggered a change in British policy. However, it is unclear why the violation of the Munich Agreement was different from Hitler's terrible behavior in the decade prior. *Why do specific violent events in the middle of power transitions trigger mistrust and competition, but not earlier violent events?*

Third, scholars are surprised by effective diplomacy. At the onset of power transitions, rising powers use diplomacy to reassure declining powers that their strategic intentions are limited even as they rapidly militarize and take territory (Yarhi-Milo, 2014; Mercer, 1996). Yet unlike crisis bargaining interactions where cheap-talk is effective (Trager, 2013, 2010; Kurizaki, 2007; Sartori, 2002, 2005; Crawford and Sobel, 1982), rising powers understate, not overstate, their intentions to avoid competition (Organski and Kugler, 1980; Vasquez, 1993). As with Hitler, even the most aggressive types claim that militarization is necessary to take a few valuable concessions, but promise that their intentions are ultimately benign. Since all rising powers understate the scope of their intentions declining powers should ignore rising powers that promise their intentions are limited (Yarhi-Milo, 2014). *Why are declining powers reassured when rising powers promise that their long-run intentions are limited?*

I offer a rational theory and formal model that resolves all of these puzzles. My key insight is that a rising power's long-term intentions depend on what motivates its foreign policy

(Moravcsik, 1998). States motivated by ethnic-nationalism, for example, covet different concessions than states motivated by security, prestige or revenge (Jackson and Morelli, 2011). For example, China would (probably) not accept territorial control over Uganda instead of Taiwan even though Uganda is larger. China's historical context and interests imply China *prefers* Taiwan over Uganda. We do not know what motivates China's foreign policy. But if we did, we would know what specific concessions China will ask for both now and as it grows stronger.

Thinking about state-types in terms of underlying motives highlights variation along two dimensions: *scope* (some want more than others); and *preference order* (some value entirely different objectives). When rising powers signal their intentions, they signal their underlying motives which provides information about both scope and preference order simultaneously.

Similar to the interaction studied by Battaglini (2002) and Trager (2011), I show that rising powers want to truthfully signal their preference order to ensure they receive valuable concessions first. Declining powers want to concede things the rising power values the most to minimize the concessions they give. Thus, both share incentives to coordinate their beliefs about preference order.

However, rising powers cannot credibly transmit information about their preference order without revealing information about scope as well. At the beginning of the power transition, the incentive to coordinate preference order overpowers the incentive to understate scope leading to credible information transmission along both dimensions. Rising powers explain what motivates their foreign policy to ensure they receive valuable concessions first. Declining powers could dismiss diplomacy as cheap-talk, or trust what the rising power says, but they prefer to evaluate if the rising power's words and deeds are consistent. As a result, they are willing to trust the rising power early in the power transition and make concessions that are consistent with the rising power's declared motives. But for the rest of the power transition, they analyze if the rising power's behavior is consistent with what was signaled at the beginning to determine if the rising power's long-run intentions are sufficiently limited.

Thus inconsistent behavior, not violent behavior, triggers mistrust and competition.

As others have shown, the incentive to understate scope is insurmountable when the rising power's value for different issues is drawn independently (Trager, 2011; Sartori, 2005), or its motives only vary in scope (Kydd, 2005). My key insight is that a rising power holds a specific foreign policy motive that causes it to care about specific territories. If rising powers could value any configuration of concessions, or their interests only varied in scope, then the mechanism I describe would not effectively rule out enough types for diplomacy to influence strategy. My mechanism only works if certain foreign policy objectives are tied to only a few different motives.

I test this argument with an elite survey experiment that simulates a National Security Council assessment of an emerging threat and randomly assigns the content of diplomatic messages and military interventions. The subjects, 93 real-world foreign policy professionals and intelligence analysts, trusted the rising power when military interventions matched the rising powers declared motives and mistrusted it otherwise. A handful of experiments demonstrate that cheap-talk affects behavior in simple laboratory settings with non-elite samples (Tingley and Walter, 2011; Cooper, DeJong, Forsythe, and Ross, 1992). 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, 2013; Mercer, 1996). Advancing a handful of studies that use elite samples (Renshon, 2015; Mintz, 2004; Hafner-Burton, LeVeck, and Victor, 2016; Friedman, Lerner, and Zeckhauser, Friedman et al.), my experiment directly observes elites in a situation that simulates their jobs.

In addition to resolving three enduring puzzles about power transitions, I also develop a new theoretical framework that connects studies about the diversity of state-motives (Mitchell and Prins, 1999; Moravcsik, 1998; Kaysen, 1990; Jackson and Morelli, 2011; Schultz and Goemans, 2015; Snyder, 1993; Schultz, 2014; Gleditsch, 2002) to broader research about conflict and cooperation (Kydd, 2005; Powell, 1999; Fearon, 1994; Mearsheimer, 2001) by

distinguishing between a state's underlying motives and the objectives it might value. My formal operationalization builds from economic models that study market actors with heterogeneous preferences (Jackson, Sonnenschein, and Xing, 2015; Battaglini, 2002). However, I constrain the type-space to reflect a difference between great powers and market actors: states have more information about their rival's history than a shop-owner does about any customer. By exploiting this contextualized information I produce more nuanced signaling dynamics that can account for qualitative differences in the territories that rising powers demand (and do not solely rely on the scope or speed of the rising power's demands).

Section 1 develops an informal theory of state-motives. Section 2 integrates my informal theory of state-motives into a formal model of power transitions. Section 3 evaluates my theory using a survey experiment with national security elites.

1 A Theory of State Motives

I define a 'state-type' as an interaction between (1) the *principles* that motivate a state's foreign policy and (2) its historical, cultural and geo-strategic *context*. In my theory, principles refer to the underlying motives that states hope to satisfy through their foreign policies (Jackson and Morelli, 2011). Scholars often theorize about one of these principles at a time. States sometimes fight to unify their ethnic group (Goemans and Schultz, 2013), restore historical borders (Carter and Goemans, 2011), for revenge (Kaysen, 1990), security (Waltz, 1979), prosperity (Keohane, 2005), status (Gilpin, 1983), regional hegemony (Mearsheimer, 2001) or the global spread of their ideology.

Holding a state's principles constant, the tangible, real-world objectives it wants to achieve depend on its historical, cultural and geo-strategic *context*. A state motivated by ethnic nationalism, for example, will be most interested in territories that contain its ethnic group. But if that same state, with the same context, was motivated by revenge from a prior conflict it would seek different objectives. Further, two states that are motivated by

the same principle will value different concessions. For example, if China and Poland both wanted to restore their historical borders their foreign policy objectives would be different because Poland's historical context is different from China's.

The combination of principles and context both drive and limit states' desire for expansion. Territorial expansion for its own sake usually costs more than what a state can extract from conquest (Brooks, 1999); especially when international trade and investment is possible (Keohane, 2005). In most cases, states will not seek territorial expansion solely for financial gain. They only pursue objectives consistent with their underlying principles. When a principle drives a state to value only a few territories, that state will exploit an opportunity to contest those high-value territories. Although many configurations of principles limit the desire to expand, others (e.g. global spread of ideology) drive states toward world domination.

To be clear, I focus on the total set of objectives that a rising power wants based on what principle motivates it. This is different from the demands a rising power makes in the short-term. For example, when Hitler first re-militarized the Rhineland he explained to the British that ethnic-nationalism motivated his foreign policy. The British understood that even if Hitler was motivated by ethnic-nationalism, that the Rhineland was not the end of his revision. They believed that Hitler's demands would encompass other Germanic territories. However, the British also believed that there was a natural limit to the concessions "ethnic-nationalist" Hitler would demand.

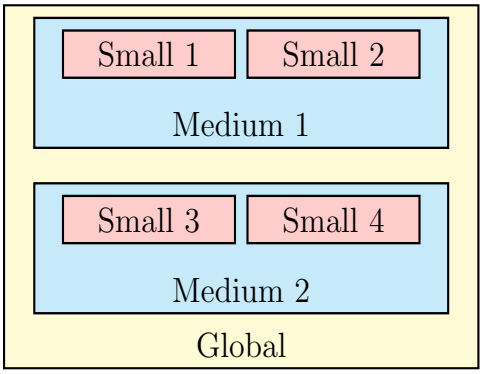
In practice, there is some ambiguity about whether a few specific issues are consistent with a rising power's principles or not. However, in most cases there is usually many more issues that clearly fit or do not fit. For example, it was clear that Africa, Asia, and the Americas did not fit into Hitler's nationalist aims but that parts of Austria, France and Czechoslovakia did.[3]

By distinguishing between the objectives that a rising power seeks and the principles that

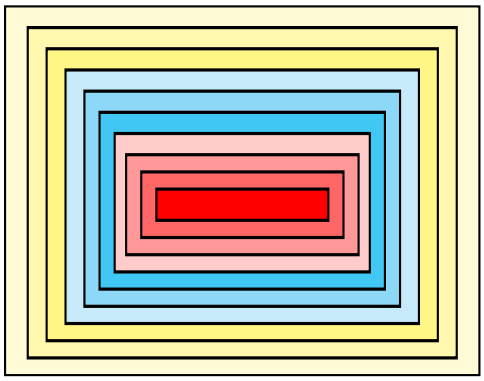
[3]Including some ambiguous issues does not change my predictions because great powers treat these ambiguous issues as if they are connected to each principle.

Figure 1: Operationalization variation in the rising power's preferences.

(a) My theory: objectives tied to principles



(b) Existing literature: Variation in scope



Each sub-Figure depicts a different type-space. Every box represents the issues that one type of rising power cares about. Colors emphasize types that care about similar numbers of objectives. Panel (a) depicts the implications of my theory of principles where there is limited variation in scope, but also variation in preference order. Panel (b) depicts the conventional wisdom where there is extensive variation in scope.

motivate its foreign policy, I assume a different set of possible rising power-types than what power transition scholars typically study. To highlight these differences, I depict a stylized version of my type-space in Figure 1(a) and the conventional wisdom in Figure 1(b). In each sub-figure, think of the total area as a map that represents all the territories a rising power can want. Each box represents a different type's core interests based on the principles that motivate its foreign policy. Each type values the issues covered by its box. The box's size represents the scope of that type's interests. The area the box covers represents the particular objectives that type values.

There are two differences between these type-spaces. First, my type-space includes an additional coordination problem: there are many rising powers that value limited objectives but each values different objectives. In standard accounts it is obvious what each type wants (and what their next demand will be) because there is only variation in the scope of the rising power's motives.

Second, standard accounts assume that there is near-continuous variation in the number of issues that a rising power can care about. This drives the logic of salami-slicing in the power transition literature: each period, rising powers make a new demand and promise this

demand will be their last (Jervis, 1978). This promise is plausible because there is always a type that values just one more concession. But this promise is not credible because types that want many more concessions will under-state their aims and pretend they want just one more concession (Mearsheimer, 2001). My type-space should exacerbate the reassurance dilemma because it is often not plausible for a rising power to want just one more territory. When a rising power makes one demand, it usually implies she wants, at minimum, several concessions.

The limited variation I describe explains why we do not observe rising powers engage in salami-slicing at opportune moments in history. For example, in 1853 president Monroe declared the United States' intention to exclude European powers from the entire western hemisphere—about 20% of the world's land-mass. Monroe did not take these territories straight away. For the next 50 years, the US made incremental demands. Under a logic or reassurance, the Monroe Doctrine was a strategic mistake because it revealed the United States held expansive aims. The US should have entered each new crisis with a new promise that this demand will be its last. My theory explains why the US was unable to revise its demands piecemeal: no underlying principle implied Monroe cared about just one issue. Britain would have been suspicious if the US claimed a single objective in 1853 absent any underlying logic to explain why that would be their last demand.[4]

1.1 British Assessment of plausible Soviet motives (1940-1942)

Archival evidence suggests that British elites evaluated Soviet post-war intentions in a way that matches my theory of principles.[5] From 1940 onward, the British War Cabinet exerted "a continual attempt to interpret Russia policy, assess the real intentions of Stalin

[4]Anglo-American relations ended in peace. But cases that eventually end in competition also start with these large claims that do not fit the logic of salami-slicing: Hitler appealed to nationalism (1932); Japan listed 21-demands to ensure their prosperity (1921); and China (1990s) demanded a resolution to 5 outstanding border disputes.

[5]This example is taken from a 70 page analytic narrative of Anglo-Soviet relations. The complete case uses over 1000 primary documents, many secondary sources and history books to illustrate my causal mechanism.

and the small governing oligarchy of the USSR."[6] However early on, as Churchill noted, a "fog of confusion and uncertainty" surrounded British beliefs about Soviet Union's post-war intentions. To remedy this uncertainty, the War Cabinet tasked several sub-Committees with studying Stalin's strategic aims.

Between 1940 and 1942, analysts made only five different arguments about what Stalin's post-war demands could have been. Consistent with my theory of principles, each argument was based on knowledge of Soviet history, and each analyst's theory about the different principles that could motivate Stalin. Advisers stationed in Moscow emphasizes that Stalin was motivated by fear from foreign invasion. Therefore, the Soviet Union would seek "the preservation of Russian interests in the Baltic and Black Seas. We may thus expect demands for Russian access to the Persian Gulf, for a revision of the Montreux Convention, possibly for the establishment of Russian bases in Norway and in Finland and the Baltic States to ensure the *security* of Leningrad and Kronstadt."[7]

Others believed Stalin's interests would converge with Tsarist Russia's historical ambitions of power politics in Asia.[8] As a result, Stalin would seek territory in Iran, the Caucuses, the Baltic States, Central Asia and, possibly, India.[9]

A report from the Foreign Office concluded that Stalin was motivated by commercial interests. They thought the Soviets "want us to approve the annexation of the Baltic States and Eastern Poland, and to help them secure special rights with regard to Finland, the Dardanelles and access to the Persian Gulf, and an ice free-port in northern Norway."[10]

Finally, a group believed that the spread of global communism motivated Russian foreign policy. Within this group, some thought Stalin wanted to expand the Soviet empire as far as he could, others thought world communism implied Stalin would subvert democracy in Asia and Europe.

[6] Woodward (1970, p105)
[7] Dew report.
[8] FO 371/248/4529 Mar 1940.
[9] COS39/66. 6 Oct. 1939
[10] FO 371/29472

Consistent with the type-space I constructed in Figure 1(a), each analyst assessed that Stalin would seek new territories once Hitler was defeated. But each analyst assessed that Stalin would seek different territories because each believed that Stalin was motivated by a different principle. For example, analysts that believed Stalin would revert to Tsarist ambitions assessed that Stalin would seek control over Iran, but analysts that believed Stalin was motivated by security or commercial interests did not. Furthermore, each analyst made radically different predictions about the scope of Stalin's demands. Those that focused on communism thought Stalin held global ambitions. But those that focused on other motives assessed that Stalin's demands would be confined to a handful of specific issues.

2 A principle-driven theory of power transitions

I build a strategic model where a rising power's motives vary because she can be motivated by different principles. There are a finite number of principles that may motivate a rising power and each drives her to value specific foreign policy objectives. When I say that declining powers are uncertain about a rising power's type, I mean that declining powers do not know what principle motivates the rising power's foreign policy. But if they did know, they would know what concessions the rising power would demand both now and in the long-run.

First, I motivate a new framework to study motives as principles during power transitions. Second, I present a simple formal model that shows complex motives facilitate trust and concessions initially, then the shift to mistrust and competition in the middle of the power transition. Third, I show that my predictions are robust to changes in my model's assumptions. Fourth, I provide an informal description of the model's informative equilibrium. Finally, I present evidence from the Anglo-Soviet case to illustrate that this equilibrium fits well with the historical record.

2.1 A motives-based framework

Recent research studies power transitions within a spatial bargaining framework (Reed, 2003; Debs and Monteiro, 2014). This research has taught us that shifting power produces incentives for incremental offers that reflect the distribution of power; and that fears of large concessions heighten incentives for competition between perfectly informed states (Powell, 1999; Krainin, 2017). But the spatial bargaining framework is poorly suited to study uncertainty about a rising power's long-term motives because short-term bargaining incentives drive rising powers to overstate, not understate their long-term intentions (cf Powell, 1999). When these models introduce incomplete information, war comes because declining powers makes low-ball offers that greedy rising powers are unwilling to accept (Reed, 2003). These dynamics may explain a few crisis episodes during a power transitions (e.g. the Alaska Boundary Dispute) but not wide-spread, strategic competition that dominates the historical record.

Older research used different assumptions from the bargaining model to study these broader competition dynamics (Powell, 1996). This research did not explicitly model power or war. Rather, it assumed that concessions moved at a fixed rate, and that states' value for competition was constant across time. I develop a framework that builds from this older literature, but also integrates insights from the bargaining model. Following Powell (1996), I assume concessions move at a fixed rate unless the rising power chooses to stop the game and accept the status quo. I define both cooperation and competition in terms of a declining power's willingness to participate in the bargaining process. The declining power *cooperates* by engaging in the bargaining process. She *competes* by refusing to make additional offers. Instead, she pays a cost to limit the number of future demands the rising power can make. Competition can take the form of major war designed to overthrow a government (a-la the Second World War), but also includes containment, and the forward deployment of forces designed to prevent the rising power from expanding (a-la the Cold War).

Following the bargaining literature, I assume that players' value for both competition

and the status quo move at the same rate. This assumption guarantees the three core result of the bargaining model obtain. First, as power shifts in the rising power's favor, she receives more concessions, and a larger expected value from competition. Second, settlements always meet both player's minimum demand from competition. Third, bargaining is more efficient than competition at every stage of the game.

Unlike either of these frameworks, I allow for variation in the order of the rising power's preferences between different issues. I expand upon Battaglini (2002), who lays out issues spatially. Building on his model, I operationalize my principle-based theory of motives by imposing a type-space similar to Figure 1(a) over spatially correlated issues.

2.2 Formal Model

I model a strategic interaction between a rising power (R) and declining power (D) that bargain over an nxn matrix Q_t^D.[11] The game models a series of signals and concessions across time. I describe an arbitrary period with a subscript t and specific periods with numeric subscripts.

Like Figure 1, Q_t^D represents a map of all the foreign policy issues that are of strategic value to D (hence the superscript D). Each element of Q_t^D can take on a dichotomous value: $\{0, 1\}$. When the element equals 1, D controls that element, when it equals 0, R controls that element. We assume that at the beginning of the game D controls all the elements, and so $Q_0^D = \mathbf{1}$, an nxn matrix where every element is 1. But since a power transition is brewing, D is forced to transfer one entry in Q_t^D to R each round.

I define a corresponding matrix Q_t^R where all element in R's controls are equal to 1, and the remainder are equal to 0. Thus, $Q_t^D + Q_t^R = \mathbf{1}$.

The following two operations make the notation simpler. First, define $A \times B = C$ as an operation between three nxn matrices. Where each entry in C is the multiplication of the

[11] To make the notation simpler, and without loss of generality, I assume $\frac{n}{4} \in \mathbb{N}$.

same entries from matrix A and B. For example:

$$\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \times \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_2 y_2 \\ x_3 y_3 & x_4 y_4 \end{bmatrix}$$

Second, define a function $f(.)$, applied to a matrix, that sums the values of the elements in the matrix together.

I assume a type space similar to Figure 1(a). I construct R's types as a series of nxn value matrices. The construction begins with the greedy type, then partitions the matrix iteratively to get variation in scope and preference order. The greedy type exists and has a value matrix **1** (every entry equals 1). I then partition the matrix into four square quarters. Each quarter is a type that has a value matrix with 1 entered into all the elements in its quarter and 0s elsewhere. The partitioning process is repeated until each type values only a single element in the matrix. The value matrices in the 4x4 game is depicted in Figure 2.

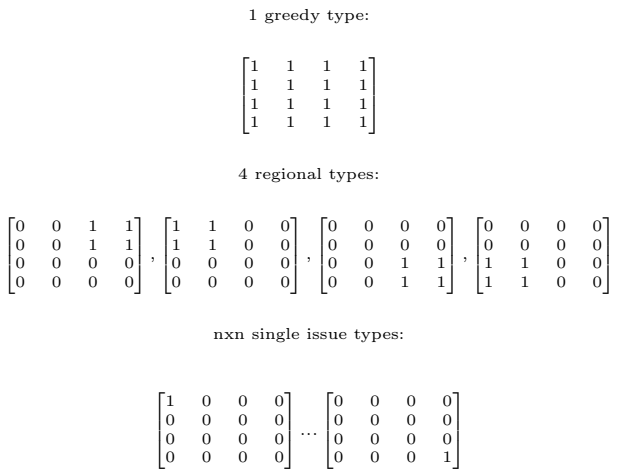
The true type V is drawn from a distribution $\Omega(w)$ over a type space $w : \{\omega_i\}$. To reflect the deep uncertainty that states have about motives during power transitions, we'll assume that each type has an equal probability of selection.

It is sometimes convenient to refer to sub-types and super-types. Any type ω_j is a sub-type (or nested type) of ω_i if all the elements that the former values so does the latter. I use set notation for sub-types and say: $\omega_j \subset \omega_i$ if $f(\omega_j \times \omega_i) = f(\omega_j) < f(\omega_i)$. For example, all are sub-types of the greedy state. Further, the single issue state that values the top left square of the matrix is a sub-type of the largest limited type that values the top left quarter of the matrix. Similarly, I refer to the inverse as super-types. A super-type contains its sub-type I sometimes refer to the next super-type as $\omega_i \supset_1 \omega_j$, and generally the m-next super-type (or $\omega_i \supset_m \omega_j$).[12]

In each period, I allow R to signal her type. R's message represents a type selected from a type-space $\Omega(w)$. I refer to the type signaled as $\hat{V}$. So R sends a message $\sigma_t(\hat{V})$ every

[12] As we shall see, costless signals eliminate all types that are not super-types.

Figure 2: Construction of types in the 4x4 game



period (t) that identifies a type $\hat{V} \in w$. R's message is honest if $\hat{V} = V$. As I explain below, D's beliefs about R's type β_t depend on the history of offers and the signals such that $\beta_t(\Omega, \sigma_t, h(\sigma, Q))$. Here β_t is a function that processes current signals and decisions not to stop the game against the prior beliefs about types and the history of signals and offers.[13]

A signal is informative when $\beta_t(\sigma_t(i)) \neq \beta_t(\sigma_t(j))$ where j includes all alternative signals. Signals do not exogenously shift the payoffs of either actor. Thus, these signals best model the effects of costless diplomacy absent costly signals such as expensive militarization choices or audience costs.[14]

At the beginning of the game, Nature determines R's type and shows it privately to R. Then the game proceeds over a series of periods where in each period:

1. R sends a message about her type, and then either decides to accept the status quo, or demand another concession.
   (a) If R accepts the status quo, the game stops and payoffs are realized.
   (b) If R demands another concession, the game continues.
2. D transfers a single element to R or decides to stop the game.

[13]Below I abuse notation slightly by writing $\hat{V}$ to refer to the type R signaled in the first period. I can do this because only the first-period signal matters.

[14]To be clear, I believe that certain costly messages may produce a similar mechanism, and even widen the conditions under which my mechanism applies. I focus on costless messages because they seem to matter in historical cases but we do not know why. I want to establish a rational foundation for that.

(a) If D transfers an element, D chooses one element to transfer and the game repeats.
(b) If D stops the game, the game ends and competition payoffs are realized.

Think of R stopping the game as if the rising power stops accepts the status quo. It may reflect a rising power diverting resources from the military to domestic spending. As a result, no more concessions are necessary and the status quo is stable, and there is no more risk of competition. If R stops the game the payoffs are:

$$U^D : f(Q_t^D) \tag{1}$$

$$U^R : f(Q_t^R \times V). \tag{2}$$

D stopping the game represents wide-scale strategic containment including forward deployed forces, an arms race, or containment policies that degrade R's access to a surplus of resources. If D stops the game payoffs are:

$$U^D : f(Q_t^D - P) - c \tag{3}$$

$$U^R : f((Q_t^R + P) \times V) - c. \tag{4}$$

Here c (a constant) is the cost of strategic competition. It is what D pays to undermine R's capacity to make future demands. P (a matrix) is a series of final concessions R can coerce under competition at her own choosing. The dynamics between c, Q and P are analogous to Soviet choices before and after containment in 1946. Prior to containment, concessions were granted as a process of negotiation between the Americans, British and the Soviets. The United States had considerable choice about what they conceded and the Soviets often accepted what they were offered. After Soviet-American competition intensified, the Soviets expanded their influence at a slower rate through proxy wars, and clandestine regime change. Although expansion was more costly post-containment, the Soviets could choose where they

made revisions unilaterally.

An equilibrium is characterized by:

- R's messaging rule $y(\sigma_t)_i$ that maximizes: $U^{R_i} : f(Q^R \times \omega_i)$ for all ω_i conditional on Ω, β_t D's offer rule O, and the stopping rules r_D, r_R.
- An offer rule (O) that generates D's order of concessions conditional on β_t.
- A stopping rule for D: $r_D | U^D(\beta_t)$.
- A stopping rule for R: $r_R | U^R(s_D, V, y())$.

2.2.1 Informative Equilibrium

The conventional wisdom is that rising powers cannot communicate information about their long-term intentions because they face insurmountable incentives to under-state the scope of their demands (Organski and Kugler, 1980). I claim there is a Pure Bayesian Equilibrium (PBE) where rising powers send an informative *costless* message in the first period to explain what principle motivates their foreign policy. This message induces early concessions, and defines the specific patterns of concessions in subsequent periods. It also produces a delayed *costly* test that declining powers will use to learn about the scope of the rising power's motives.

From D's perspective, my model produces the same strategic incentives as in standard accounts: D only cares about the total number of concessions he will be forced to make across the game. To properly characterize D's incentives define:

- a threshold T_{ω_i} as the fewest concessions that will completely satisfy type ω_i.
- a probability $\lambda_{Tj} : pr(r_R = T_{\omega_i} | \beta_t(\sigma), O, h(O, \sigma))$ that R is a type that can be fully satisfied following T_{ω_i} concessions.

The subscript T on λ refers to the threshold T_i implied by the signal received $\sigma(\hat{V})$. The subscript j refers to the time period (t) that this signal was first made. For example, a first round signal of the most greedy type in a game where Q is an 8x8 matrix is $\lambda_{T=64,j=1}$.

The key to the result is a critical threshold: T_1^*; and corresponding message $y(\sigma(\hat{V} \leq T_1^*))$. The subscript 1 on T_1^* emphasizes that this threshold emerges from a first round signal. The integer T_1^* represents the maximum number of concessions that D prefers to make in the first round instead of enacting competition (conditional on λ_{T1}). By definition of T_1^*, if R signals some type who values T_1^* elements in the matrix or fewer, D prefers to make concessions in the first round and subsequent rounds if D's beliefs remain constant.[15]

I focus on messaging strategy $y(\sigma(\hat{V} \leq T_1^*))$. In it, R observes her type then in the first period signals honestly ($\sigma_1(\hat{V}) = \sigma_1(V)$) if R values T_1^* or fewer concessions ($f(V) \leq T_1^*$). If R values more than T_1^* concessions R signals dishonestly in the first period. These dishonest messages have two features. First, R understates the scope of her demands by claiming that she holds a principle that values exactly T_1^* concessions. Second, R's message always reveals a type that is a sub-type of R's true type ($\hat{V} \subseteq V$.). Even though R does not reveal her true aims, she sends a message that reveals T_1^* issues she actually cares about. Types who signaled dishonestly in the first period all send a new message in period $t = T_1^*$.[16] They all promise that they are the next-largest type. For a technical description of this message see Appendix A.1.

We are now ready to characterize behavior in the informative equilibrium. First, I define the equilibrium conditions in a proposition. I then prove these behaviors form an equilibrium under the assumption that the critical threshold T_1^* exists. Finally, we show that T_1^* exists so long as $\frac{T_{\omega_i}}{P+c} - \lambda_{1j} > 0$ can be solved for some T_1.

Proposition 2.1 *If $\frac{T_{\omega_i}}{P+c} \geq \lambda_{T1}$ can be solved for some arbitrary first round signal $\sigma_1(\omega_i)$ then, a critical threshold T_1^* emerges that implies the following set of strategies form a PBE:*

- *R observes her type, V, then:*
  - *If $f(V) \leq T_1^*$ R signals honestly $\sigma_1(\hat{V}) = \sigma_1(\hat{V})$ in the first period, and repeats this signal every period.*
  - *If $f(V) > T_1^*$ R signals dishonestly $\sigma_1(\hat{V})$ in the first period such that $f(\hat{V}) = T_1^* < f(V)$. Further, the signaled type $\hat{V}$ is a randomly selected sub-type of the*

[15] To be clear, we do not assume T_1^* exists. We show that it does.

[16] I write $t = T_1^*$ to make explicit I am referring to a specific period, and not T_1^* concessions.

true type such that $\hat{V} \subseteq V$. R repeats this message for T_1^* periods, then signals the next largest type in which $\hat{V}$ is nested.

- *D processes R's message in one of two ways:*
  - *If R's first period message implies: $f(\hat{V}) < T_1^*$, then D adopts beliefs $\beta_t|\sigma_1(\hat{V}) \implies pr(V = \hat{V}) = 1$.*
  - *If R's first period message implies: $f(\hat{V}) = T_1^*$ then D adopts beliefs that R's true type is either $\hat{V}$ or some super-type of $\hat{V}$. D's beliefs are structured such that $pr(V = \hat{V}) > pr(V = \omega_j \supset \hat{V})$ for all super-types, ω_j.*
- *D offers elements equal to 1 in $\hat{V}$ in random order.*
- *R stops the game once $Q_t^R = V$ if D has not.*
- *D stops the game at $\hat{T} + 1$ if R has not. When $f(\hat{V}) = T_1^*$, D stops the game after exactly $T_1^* + 1$ rounds.*

We assume that if D observes off-path behavior, D believes that R is the type that maximally profits from this deviation.

I report the complete solution in Appendix A.2 and a numeric example in Appendix A.3. In this equilibrium, if R's first period message implies she cares about T_1^* or less issues, D will concede the territories that R claims to value. To start, I'll assume that the critical threshold T_1^* exists and consider R's incentives to send different first period messages.

Rising powers that value T_1^* or less issues ($f(V) \leq T_1^*$) can send an honest message and receive all the issues they value. Clearly, they cannot profit by sending a different message.

Types that value more than T_1^* issues ($f(V) > T_1^*$) face competition if they signal honestly. Wanting to avoid competition and maximize the concessions they receive, these types pool on a signal of size T_1^*. By definition of T_1^*, this guarantees that R will receive exactly T_1^* concessions before competition is enacted. But even these types with expansive aims want to ensure the concessions they receive are valuable, and so signal some type nested within their true type.

It follows that if D observes a first period message $\sigma_1(\hat{V}) = T < T_1^*$ that D is certain that the message is honest. However, if D observes a first period message $\sigma_1(\hat{V}) = T_1^*$, then D is uncertain if R's message is honest and R actually values T_1^* issues, or if R is really a type $f(V) > T_1^*$ who is understating her motives.

Nevertheless, when D observes a first period message $f(\hat{V}) = T_1^*$, D updates his beliefs for two reasons. These reasons emerge because R's motives vary in preference order (that is, there are different types that each value exactly T_1^* concessions). First, D rules out the possibility that R is any type other than $\hat{V}$ and all super-types of $\hat{V}$.

Second, of these remaining types, each has a different probability of sending the message $\sigma_1(\hat{V})$. As discussed above, the type who truly values T_1^* issues always sends an honest messages. Since this type always signals honestly, D's posterior belief that R's first period message is honest is:

$$\beta_1(V = \hat{V}|\sigma_1(\hat{V})) = \frac{P(\hat{V})}{P(V = \hat{V}|\Omega, y) + P(V \neq \hat{V}|\Omega, y)} \tag{5}$$

In contrast, types with more expansive aims will under-state their aims. But since they value more than T_1^* issues, they can find many sub-types that value exactly T_1^* issues. These greedier types choose from one of the many different lies that they can tell that will produce T_1^* valuable concessions. It follows that if D observes a first period message $f(\hat{V}) = T_1^*$, D's beliefs that this message is dishonest, and R's true type is the m-next largest super-type of $\hat{V}$ is:

$$\beta_1(V = \omega_{+m}|\sigma_1(\hat{V})) = \frac{P(\hat{V})}{4^m * (P(V = \hat{V}|\Omega, y) + P(V \neq \hat{V}|\Omega, y))} \tag{6}$$

Even though D initially believed that each type was equally likely, D down-weights the possibility that R is dishonest because greedier types mix over the different messages they can send which imply they care about T_1^* issues. Taking equation 18 as a sequence and summing the probabilities that R is a type that values more than T_1^* issues, $\lambda_{T^*1} \to 75\%$ (from above) as $nxn \to \infty$.

We are now ready to consider D's first period expected benefit from playing this equilibrium strategy. In the first period, suppose D observes a message that implies R wants T_1^* concessions. Then, D prefers to make concessions for T_1^* periods and compete in period

$T_1^* + 1$ if he discovers that R was initially dishonest if:

$$\overbrace{(N - T_1^*)}^{\substack{\text{D's value if R stops} \\ \text{at } t = T_1^*}} \underbrace{(1 - \lambda_{T1})}_{\substack{\text{pr. R stops} \\ \text{at } t = T_1^*}} + \overbrace{(N - T_1^* - c - f(P))}^{\substack{\text{D's value if D stops} \\ \text{at } t = T_1^*}} \underbrace{\lambda_{T1}}_{\substack{\text{pr. R doesn't} \\ \text{stop at } t = T_1^*}} > \overbrace{(N - T_1^* - c - f(P))}^{\substack{\text{D's value for stop-} \\ \text{ping straight away}}} \quad (7)$$

$$\equiv \lambda_{T1} > \frac{T_1^*}{f(P) + c} \quad (8)$$

λ_{T1} is D's belief that R will stop following T_1^* concessions given R's first period message $\sigma_1(\hat{V})$, which implied that R valued T_1^* issues. Inequality 8 is always satisfied if we can find a T_1^* such that $\frac{T_1^*}{f(P)+c} < 0.75$.

Inequality 7 assumes that D can credibly promise to select competition at round $t = T_1^*$, if D discovers that R's original signal $\sigma(\hat{V})$ was dishonest. In theory, it is possible that R can send a second informative message at T_1^* and convince D to make additional concessions. In Appendix A.2, I show that R's second message is never informative. The reason is that in period $t = T_1^*$ all types who keep making demands promise to be the next largest super-type of $\hat{V}$. As a result, their second message is not informative. and D switches to competition.

2.3 Model Robustness

Some argue that power variables are more important for explaining competition during power transitions than changes in the declining power's shifting beliefs about the rising power's motives. Others argue that great powers start out confident that rising powers are the greediest possible type, or that rising powers can start out with limited aims and become greedy over time (Mearsheimer, 2001).

In Appendix B I analyze extensions to the model to address these alternative theories. First, I introduce shifting power and give both players an outside option of war in the form of a costly lottery (a-la Powell, 1999). Second, I assume that R values peripheral interests $L > 0$, and not 0. Third, I assume that R can start out with limited aims but become greedy

with probability. Fourth, I assume D starts out with stronger priors that R is the greediest possible type. In every extension, I find an equilibrium that is substantively the same as the informative equilibrium described in proposition 2.1.

These extensions help set important scope conditions for my theory. When power shifts rapidly, or the declining power holds prior beliefs that the rising power is likely to be (or become) greedy, or the rising power cares about peripheral interests almost as much as her core interests, the declining power enacts competition following the rising power's first demand; and diplomacy has no reassuring effect.[17] However, in moderate parameter ranges, my motives-based logic dominates the strategic setting in every extension.[18] Thus, I consider my theory and theories of shifting power to be complimentary mechanisms that are important given different prior beliefs, competition parameters and rates of shifting power.

2.4 Informal Description of Equilibrium

In the informative equilibrium a critical threshold emerges that defines the number of concessions the declining power is willing to make before enacting competition. We write that number as T_1^* concessions. We use subscript 1 to emphasize that the threshold emerges in the first round but governs concessions over the entire power transition.

This threshold depends on the declining power's cost for competition, and prior beliefs about the rising power's intentions. The threshold allows me to endogenously define limited aims and greedy rising powers in terms of the amount of revision that the declining power is willing to accept at the onset of the power transition: Greedy types value more than T_1^* concessions; acceptably limited aims types value T_1^* or fewer concessions.

To start, I summarize the strategy and beliefs for both players based on this critical threshold. At the onset of the power transition the declining power process the rising power's message and makes concessions that are consistent with that message. If the rising power

[17] Cases that fit this logic include Franco-Prussian (1886) and Sino-Soviet bargaining.

[18] Cases that fit this logic include Anglo-German (1932-1939), Sino-American (1990-present), Anglo-American (1850-1910), Japanese-American (1980s), German-American (1990) bargaining.

signals that she values fewer than T_1^* concessions, the declining power believes that her signal was genuine and concedes exactly what the rising power claims to value.

If the rising power signals a type that values exactly T_1^* concessions, the declining power becomes sufficiently optimistic that the rising power's aims are limited and selects cooperation over competition. Before the signal, it was unclear what particular concessions were valuable to the rising power. Afterwards, the declining power believes that if the rising power has limited aims she wants what she said she wants.[19] During the power transition, the declining power cooperates for T_1^* rounds, so long as the rising power keeps sending the same signal. If the rising power ever alters her signal, or does not stop the game at round T_1^*, the declining power infers that the signal was dishonest. At this point the declining power shifts its beliefs. Although it does not know precisely what the rising power wants, it knows that it is more than what the declining power is willing to give. If the declining power observes a different message after the first period, the declining power enacts competition.

All rising powers want to avoid competition for as long as possible. As a result, they all reassure the declining power that their intentions are limited. Those with limited aims will be honest. They will signal their true type, and make territorial demands to capture the concessions they value. But once they capture all the concessions that they value they will stop.

Greedy types will pretend to hold limited aims to avoid competition. Since they will receive the concessions that are consistent with their signal, they promise that their intentions are limited in a way that ensures they still receive valuable concessions. Furthermore, they maximize the number of concessions they receive by signaling a type that is right at the threshold. Once greedy types have received T_1^* concessions, they are forced to reveal their true type and face competition.

In what follows I explain the two most surprising features of this result. First, the rising power's first period message is so informative that the declining power is willing to concede

[19] To be clear, the declining power still believes that the rising power could be a greedier type.

everything that the rising power claims to value, rather than compete in the first period. This is true even though the declining power knows that greedy rising powers will understate their motives. Second, once the declining power discovers that the rising power's first period message was dishonest, the declining power switches to competition. The rising power is unable to send a second reassuring message. This is surprising because the rising power's first period message was reassuring. If diplomacy worked once, why can't it work a second time?

2.4.1 Credible Diplomacy at the Onset of Power Transitions

At the onset of the power transition, greedy and limited aims rising powers face different information problems because they do not care equally about resolving uncertainty about the scope and preference order of their motives. Limited aims rising powers have no incentive to understate their true intentions because they can truthfully reveal the scope of their interests in the first period and still avoid competition. But in my theory, rising powers care about specific concessions because I explicitly allow for variation in preference order. Thus, if limited aims rising powers do not reveal which concessions they value, they will not receive valuable concessions. Since limited aims rising powers value only a few, specific issues they have strong incentives to coordinate with the declining power about their preference order.

In contrast, greedy rising powers want to maximize the number of concessions that they receive before they face strategic competition. However, they will face competition in the first period if they truthfully reveal their motives. As a result, greedy types prioritize the incentive to understate their intentions and signal that their core interests are less than what they really are. Of course, greedy types get to choose the signal that they send. Since they want to maximize the number of concessions they receive, they dishonestly signal that they are a type that cares about exactly T_1^*.

It follows that when D observes R send a message that implies she values exactly T_1^* issues, D does not know if R is honest, or if R is a greedier type who under-stated her aims.

But D can exploit R's incentive to coordinate on preference order to make three inferences about R's type. First, D can rule out all the limited aims types except the one that was signaled because each limited type signals truthfully.

Second, D can rule out all greedy types that place no value on what was signaled. The reason is that R expects concessions that are consistent with the type she signals. All greedy types lie about their interests to avoid competition. But they lie in ways that assure they receive valuable concessions. Thus, they signal a smaller type nested within themselves. As a result, only the type that was signaled and the greedier types that value the same concessions remain feasible options.

Third, D believes that the signaled type is more likely to be true than any of the greedier types *even given R's incentive to understate her scope*. The reason is that greedy types have several different messages to choose from that will convince D to concede T_1^* concessions to them. A rising power that wants world domination can say she wants to restore its borders, or security or to unite its ethnic kin. Because it values all concessions implied by all these principles, it is indifferent about which message it sends. Similarly, a rising power that wants regional hegemony can select the lie it tells from all the subtypes that value concessions within its region of interest.

Rational updating discounts all the greedier types because there are many lies they can tell and, therefore, *just some probability of telling a specific lie*. Only the limited aims type that truly values T_1^* issues would have told the truth with certainty. Bayes' rule factors in the probability that the declining power observes a specific message, conditional on the probability that each type would have sent it. The resulting distribution is skewed in favor of thinking that R's first period message was honest because only the honest type sends that message with 100% probability. All greedier types who are understating their motives had some chance that they would have sent a different message.

Given that greedier types send mixed signals, D is sufficiently confident that R's first period message is honest that he is willing to make exactly T_1^* concessions that are consistent

with R's original message. D is willing to do this because he knows that he will learn in period T_1^* whether R's original message was honest or not, and is willing to wait that long to enact competition given his expectation that R's initial message was honest.

Hypothesis 1: *A declining power will increase its confidence that it understands what a rising power's intentions are after the rising power communicates it's intentions for the first time.*

2.4.2 Why Inconsistent Behavior Triggers Alarm Bells

After receiving the initial message, the declining power makes concessions to satisfy the type that R signaled. Clearly, there will be a point where no more concessions remain that are consistent with that signal. At that point, the rising power faces a critical juncture: stop making demands[20] or send a different message about her motives. One might expect that R could update her message as her bargaining position improved. After all, if cheap reassurances worked at the onset of power transitions, then why can't the rising power just say it wants a little bit more when she is stronger?

It turns out that revised signals do not have the same reassuring effects that the original signal had. The reason is that the original signal eliminated every type whose interests did not intersect with the type that was signaled. No additional signal can rule out these types again. As a result, once D observes R's first message, D can eliminate uncertainty about R preference order. Now only variation in scope remains. We know from the existing literature that R cannot credibly under-state her motives when her motives only vary in scope. The reason is that all types want to promise that they are the next largest type. Since all remaining types will send the same message with 100% probability, D is not re-assured by R's claims that she is the next largest type. The result implies that diplomacy works well the first time, but cannot be used to set a second threshold.

The two stages of the updating process were described to me during elite interviews. The

[20]Examples of limited types that chose to stop making demands include Japan (1980s) and Germany (2000s).

way that early diplomatic messages form a benchmark was well summarized by the former Ambassador to Pakistan who now works on China issues. During the interview, I explained the problem of cheap-talk to the Ambassador. I then asked him:

> **Researcher:** When you are dealing with counterparts and you believe they may have an incentive to misrepresent, how can you trust what they tell you?
>
> **Amb. Hayward** Well you don't just believe what they say. But you need a basis against which to build a relationship. Once they tell you what they want, you have that basis. Then when something seems different you can say: "hang on, you told me this six months ago."

The updating process was explained to me by the former Director of National Intelligence Dennis Blair when I interviewed him about his assessment of Chinese intentions:[21]

> **Researcher:** Up until the middle of 2010 you call upon U.S. policymakers to look for opportunities to cooperate with China. From 2010 onward that call is stripped from all of your writings and public statements. What changed? Was it access to different intelligence or something else?
>
> **Blair:** You look at the harder line that they [China] have taken in the South China Sea since 2010. They are pushing pretty hard to take control of that area. Pushing the U.S. out. It was out of character for what they'd been doing up until that point... It was opportunistic.
>
> **Researcher:** As a hypothetical — if they had behaved the same way against Taiwan... if they ran the same bellicose and aggressive missions around the Strait of Taiwan would you have updated your assessment in the same way?
>
> **Blair:** No I don't think it would have been the same. That would be different. I would have seen that as speeding up the time-table on something they had long-declared that they had been serious about. I think it has been the taking on a new area that was not mentioned as a core interest that influenced the assessment.

This leads to two additional hypotheses that relate the original claims to behavior during the power transition:

[21] The DNI is the most senior intelligence official in the United States.

Hypothesis 2: *Declining powers increase their confidence that a rising power's intentions are greedy after they witness behavior that is inconsistent with its history of diplomatic messages.*

Hypothesis 3: *Declining powers that observe inconsistent behavior are less likely to trust what rising powers say than if they observed consistent diplomacy and behavior.*

2.5 Illustration: British assessments of Soviet intentions (1940-1946)

I report archival evidence from Anglo-Soviet (1940-1946) relations to demonstrate that my mechanisms matches how the British formed and updated their beliefs about Soviet intentions. I focus my analysis on two critical junctures: when Stalin first justifies his long-term strategic motives during a diplomatic meeting with Anthony Eden (1941); and the Iran Crisis (1946), when British elites realize that Stalin's initial diplomatic justification was dishonest.

In section 1.1 I argued that the British produced five theories about Stalin's possible motives in 1940; each grounded in a different principle. In this section I ask: how confident were the British in each of these assessments in 1940; what events triggered the British to update their assessments?

In 1940, British elites were uncertain, not pessimistic, about Soviet intentions. In all five assessments described in section 1.1, analysts reported low confidence in their own assessment because "the Russians have been extremely reticent in defining their ideas on war aims and the post war settlement."[22] Recognizing that an explanation from Stalin was a critical piece of evidence, Foreign Minister Eden was sent to Moscow to resolve this uncertainty in November 1941. Before he left, Eden realized that "we ought to be examining the question of our post-war relations with the Soviet Government as far as it is possible to do so at

[22]FO371/29472, N5679/3014/38

the present stage of the war."[23] To that end, he had given the subject of Soviet intentions considerable study before embarking on his journey. Yet on the eve of his visit to Moscow he still was uncertain about Soviet intentions because he had "not received and explanation from Stalin as to what he himself has in mind when he proposes a post-war alliance."[24]

On the first day of Eden's visit to Moscow Stalin blurted out his post-war interests. By Eden's recollection, "At my first conversation with M. Stalin and M. Molotov... Stalin set out in some detail what he considered should be the post-war territorial frontiers in Europe."[25] Eden then recounted Stalin's detailed list of demands which included the dismemberment of Germany, Soviet control of Polish territory up to the Curzon line, and control over Baltic states, Finland and Bessarabia.

Stalin's demands were enormous. He asked for territory from six sovereign states, expand military bases through Europe and Asia and permanently dismember Germany—the only counterbalance on the continent. His statements left no doubt: the Soviet Union had revisionist intentions.

Stalin did not just make demands, he justified them: "We must have these [frontiers] for our security and safety." The next day Eden made clear that he "fully realize[d] that you [Stalin] want security on your north-western frontier."[26]

Following Eden's visit, there was a clear change in how British analysts evaluated Soviet intentions. Most notably, security entered the debate as the focal point for assessing Stalin's post-war aims. Many policy-makers who before Eden's visit were reserved, now mentioned security explicitly as Stalin's most likely motivation. Even many of those who began mistrustful of Soviet post war aims now described Soviet behavior as in service of "their own security after the war."[27] Upon reading Eden's report, the War Cabinet tasked sub-committees to establish if Stalin truly was motivated by security.[28]

[23]Prem 3/395/6, Nov. 10, 1941
[24]Prem 3/395/6, Nov. 10, 1941
[25]Prem 3/394/3 Dec. 18, 1941
[26]WP(42)8; N109/5/38(1842).
[27]FO 371/32/876 Feb. 12, 1942
[28]See WP(42)69 ; WM(42)18 ; N798/5/38 ; N1024/5/38; N1279/5/38; N1526/5/38; T395/2/402

This evidence supports my first hypothesis about why private diplomacy is so important for coordinating beliefs early on. The British started out uncertain about which principle motivated Stalin's policy. They explicitly recognized that their uncertainty was so high because Stalin had not explained his interests. Thus, they recognized that diplomacy played an important role in coordinating their beliefs. Before Stalin explained his motives, British elites were most interested in developing theories about what Stalin could want. After Stalin explained his motives, the British changed their focus. They started to analyze whether or not Stalin was actually motivated by security as he claimed.

The most direct evidence for my mechanism comes from Eden's explanation for why he found Stalin's diplomatic justifications re-assuring:

> "It must be remembered that Stalin might have asked for much more, e.g. control of the Dardanelles, spheres of influence in the Balkans, one-sided imposition on Poland of Russo-Polish frontier, access to Persian Gulf, access to Atlantic involving accession of Norwegian territory. Stalin's present demand it is true, may not be final, but he may later be in a position to enforce a claim to some or all of these, and we and United States Government would be in stronger position to assert our views if we have established precedent of tripartite agreements in regard to post-war arrangements, and if Soviet Government have not decided to go ahead without regard to our views owing to our giving an entirely negative reply to present demands."

Eden recognized that Stalin faced incentives to understate his aims. Yet Eden found Stalin's claims credible because Stalin could have suggested a different principle and asked for different concessions. The fact that Stalin chose to signal one principle, restricting what he could claim in the future, made the statement more credible.

(Churchill Papers); N1395/5/38); N1653/5/38; WM(42)37, C.A. for extensive discussion about Stalin's intentions relating to security.

2.5.1 Why the British came to mistrust Stalin

Throughout the Second World War there was a constant stream of events that should have altered British assessments of Soviet intentions according to existing theories. Notably, there were eight in-person meetings between the heads of state, and or foreign minister, three treaties signed, several political spats and ominous military interventions. Towards the end of the war the Soviets and the Western allied explicitly bargained over the fate of Europe. During that time, the Soviet Union made clear its interests in annexing territory. Indeed, if cognitive biases brought on during personal interactions (Yarhi-Milo, 2014), violent military behavior (Waltz, 1979), taking territory (Glaser, 2010), shifts in the rising power's domestic political discourse, cause declining powers to update their assessments, I should observe British assessments shift frequently during the Second World War, and following the Conferences at Yalta and Potsdam, and elsewhere.

Yet British assessments of Soviet intentions did not change during this period. Instead, the British focused on whether or not the Soviets were truly motivated by security. This was clear in a December 1944 report by the Joint Intelligence Committee, the peak war-time intelligence body titled, "Russia's Strategic Interests and Intentions from the Point of View of Her Security." The "hugely detailed"[29] report was approved by the Foreign Office and the Chief of Staff Committee (COC) and sent to the War Cabinet and prime minister.[30]

The JIC judged that at minimum, "in order to achieve the greatest possible security Russia will wish to improve her strategic frontiers and to draw the States lying along her borders, and particularly those in Europe, into her strategic system. Provided that the other Great Powers are prepared to accept Russia's predominance in these border States and provided that they follow a policy designed to prevent any revival of German and Japanese military power, Russia will have achieved the greatest possible measure of security

[29] Goodman (2014)

[30] Four similar reports were released. The fact that these reports ever emerged supports my theory. No reports, emerged that looked at what the Soviet Union would want if it was motivated by nationalism. See Goodman (2014); Aldrich, Cormac, and Goodman (2014).

and could not hope to increase it by further territorial expansion. Nor is it easy to see what else Russia could under such conditions hope to gain from a policy of aggression." The report identified that "Russia would regard Finland, Poland, Czechoslovakia, Hungary, Romania, Bulgaria and to a lesser extent Yugoslavia as forming her protective screen. She will, however, probably regard Norway and Greece as being outside her sphere."

In the months following Germany's surrender, aggressive Soviet behavior put this framework to the test. For example, two months after the Yalta Agreement, Stalin orchestrated a coup in Romania and installed a communist government in Poland.[31] Observing these events, a small alarmist group argued, "the fact is that the Soviet Government have not hitherto been given any compelling reason to suppose that that we should insist on hit moderating its ambitions and behavior."[32]

Despite these claims, most British policy-makers called for calm. While Soviet behavior concerned them, they argued that the Soviets "intend, however, to secure their own essential interests, and in particular to buttress the Russian frontiers against any possible renewal of German aggression..."[33] Kerr acknowledged that Soviet behavior was underhanded and opportunistic "But this should not, I think, be interpreted as a sign of hostility to the west or as a danger signal for the future"[34] because "Russia policy, however distasteful it may be to us... has the air of remaining a policy of limited objectives... in the case of Greece, they have refrained from intervention and shown what is for them extreme moderation... [In Persia] they have in fact refrained from reviving their demands for oil concession and they seem to have realized that the independence of Persia of a matter of vital importance."[35] These calls for calm persuaded the balance of policy-makers including the Prime Minister and Foreign Minister.[36]

The balance of policy-makers altered their opinion after the Iran Crisis in February

[31]He promised not to do this at Yalta.
[32]FO/371/50912 Lockhart's Minute, Apr. 11 1945
[33]Clark Kerr's opinion: FO 371/47076 Apr. 16, 1945
[34]Clark Kerr's opinion: FO 371/47076 Apr. 16, 1945
[35]Clark Kerr's assessment: FO371/47941 Mar. 27, 1945
[36]who concurred on this Memorandum.

1946. Following these events most policy-makers, including the Prime Minister and Foreign Minister, changed their assessment of Soviet intentions. The Iran Crisis was no more violent than Soviet invasions that came before. Yet it drove many analysts to alter their assessment. The JIC made explicit why the Iran Crisis was unique. "Our report of the 18th December, 1944,... [concluded that] Russia's policy after the war would be directed primarily towards achieving the greatest possible measure of security... [We now believed that] Russia will seek, by all the above means, short of major war, to include within her "belt" further areas which she considers it strategically necessary... [S]he will adopt a policy of opportunism to extend her influence wherever possible."[37] The JIC altered their assessment because the Iran Crisis demonstrated that "The Soviet Government have, in fact, resumed the traditional Russian policy of southern expansion, which was temporarily suspended between the fall of the Czarist regime and the war of 1939... She will seek to extend her influence wherever possible in the world."[38]

The evidence supports my final two predictions. In the face of violent behavior, several diplomatic meetings and territorial demands, British assessments of Soviet intentions remain constant because Soviet behavior was consistent with Stalin's stated aims. Only once Stalin demanded territory beyond what was plausibly in the service of security did the British alter their assessment. As the JIC assessments confirm, the reason their assessment changes is because Soviet behavior is inconsistent with their stated aims.

3 Causal Evidence

I report evidence from an elite survey experiment on 93 foreign policy, intelligence and defense professionals. The vignette presented subjects with information about a fictional rising power named Bandaria. Subjects were randomly assigned information about Bandaria's diplomatic messages and intervention choices then asked to assess Bandaria's intentions. The

[37]Report of the Joint Intelligence Sub-Committee of Great Britain: 'Russia's Strategic Interests and Intentions,' Mar. 1, 1946.

[38]JIC(46)38(0) Final, Jun. 14, 1946. 'Russia's Strategic Interests and intentions in the Middle East.'

hypotheses and research design were pre-registered at [removed-for-review].

Although the historical evidence is supportive, the rising power's diplomacy and military behavior is endogenous to a strategic process. Power transitions last for decades and it is difficult to know if the statements made well capture the true logic of the core decision-makers. Random assignment allows me to overcome much of these concerns and isolate the effect of consistent words and deeds.[39]

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.[40] 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

[39] Experiments have limits. To address these limits I include archival evidence, and cross-national statistical analyses in my book manuscript to support my theory.

[40] 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.

Research shows that professionals with specialized expertise approach tasks associated with their work differently than the average educated adult. 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 (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 3. 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 11 reports the summary statistics for my sample and the real NSC staff. It shows that biographical features in my sample are consistent with NSC staff. 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, rising powers are not simply profit maximizers that respond to well defined purchasing choices (like individuals in markets). There are several dimensions of preferences and outside options that may effect decision-making. 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[41] and National Security Estimates[42] 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 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 concerns 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 features, attention checks, and meta-data that demonstrate the subjects took the vignette

[41] That is, the US military's specific plans to invade other countries

[42] That is, the intelligence community's assessments of foreign threats.

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

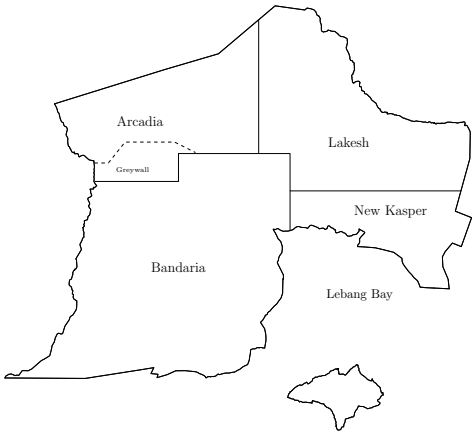
The survey instrument, presented in Appendix C, has three phases. In each phase, subjects are presented with new information about Bandaria then asked standardized questions about Bandaria's long-term intentions. Subjects also write text responses.

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 in preparation for that meeting. The baseline vignette 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 3. 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

Figure 3: Bandaria and its Neighbors



survey suggests that the level of detail was necessary to make the scenario realistic.

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 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 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 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.

I modeled this language on de-classified minutes, letters and cables that described conversations between British elites and German Kaiser Wilhelm (1866), US president McKinnely (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 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.[43]

Although my theory applies to inconsistent words and deeds broadly, I focus on milita-

[43] After manipulation and selection checks, consistent = 29, inconsistent = 31, control = 33.

rized interventions for two reasons. First, conflict parses my theory from the main alternative explanations. Existing rationalist scholarship argues that military spending and interventions signal aggressive intentions and should raise alarm (Waltz, 1979). Existing psychological explanations argue that analysts update based on impressions formed during meetings (Yarhi-Milo, 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. Second, conflict addresses desirability bias. Since conflict is widely thought to signal aggressive strategic intentions, subjects are likely to update based on observing any violent military event, no matter the diplomatic context. This bias should lead all subjects to update their beliefs. Given these factors, it would be unusual to observe differential effects between consistent and inconsistent military behavior, or to find no significant difference between the consistent group's response before and after they observe conflict.

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

3.3 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.[44]

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.[45]

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:

$$e \cup s > c \tag{9}$$

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:

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

[45]The 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 improves my interpretation of any result reported in the paper.

> **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.

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:

$$eS \cup sE > eE \cup sS. \tag{10}$$

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:

$$(eS \cup sE)_3 - (eS \cup sE)_2 > (sS \cup eE)_3 - (sS \cup eE)_2. \tag{11}$$

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: $(sS \cup eE)_3 - (sS \cup eE)_2 \approx 0$.[46]

[46] I present alternative comparisons to all these tests between consistent message and the control groups

3.4 Results

Figure 4 presents results for hypothesis 1. The top of Figure 4 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 4 plot responses to question **A**. The lighter mass received a diplomatic treatment that contained a reassurance (group mean is the solid line). The darker 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.[47] 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 5 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

in Appendix E. The results are substantively same.

[47] Although the distributions are shaped differently the means are the same.

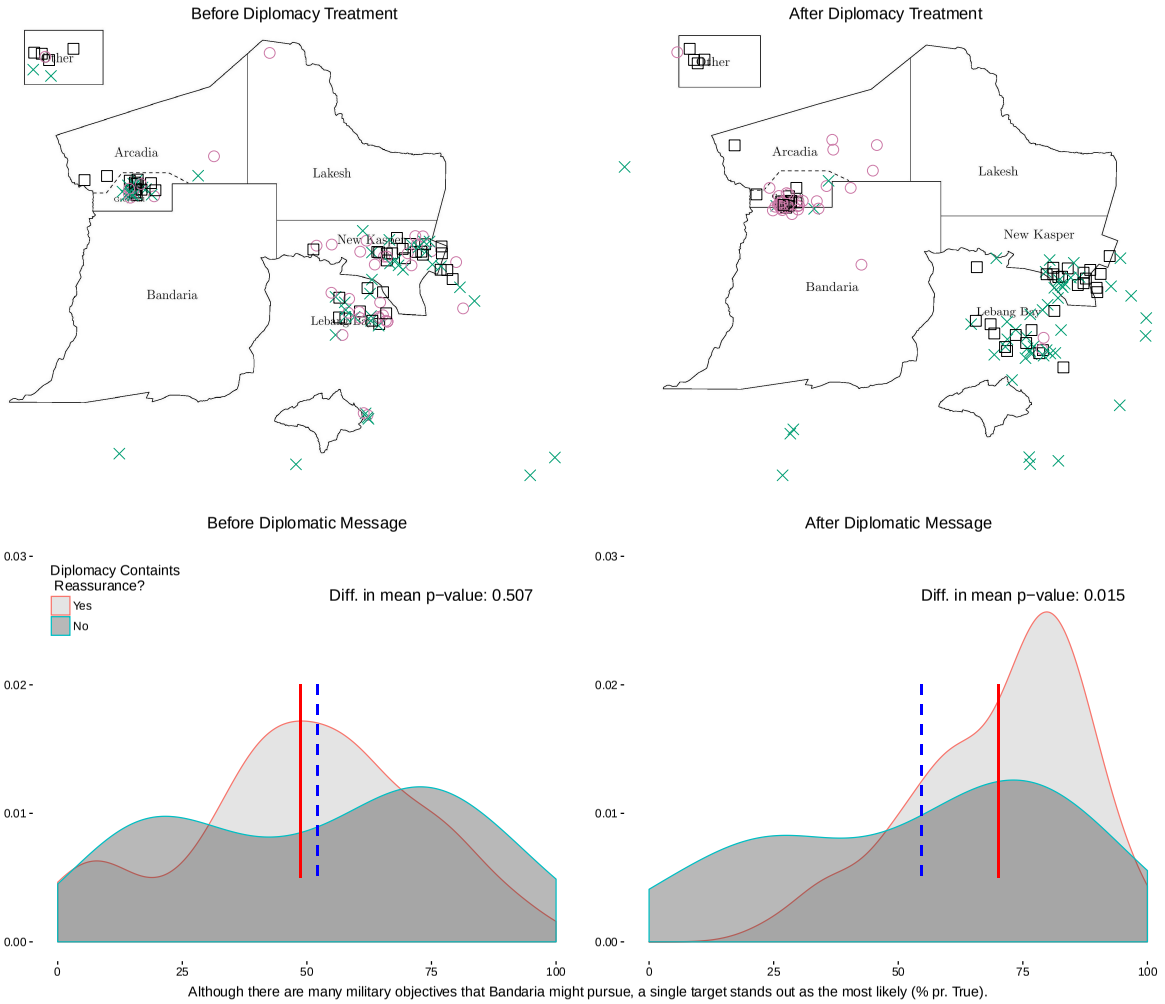


Figure 4: Does Cheap Diplomacy Influence Beliefs About Limited Aims?

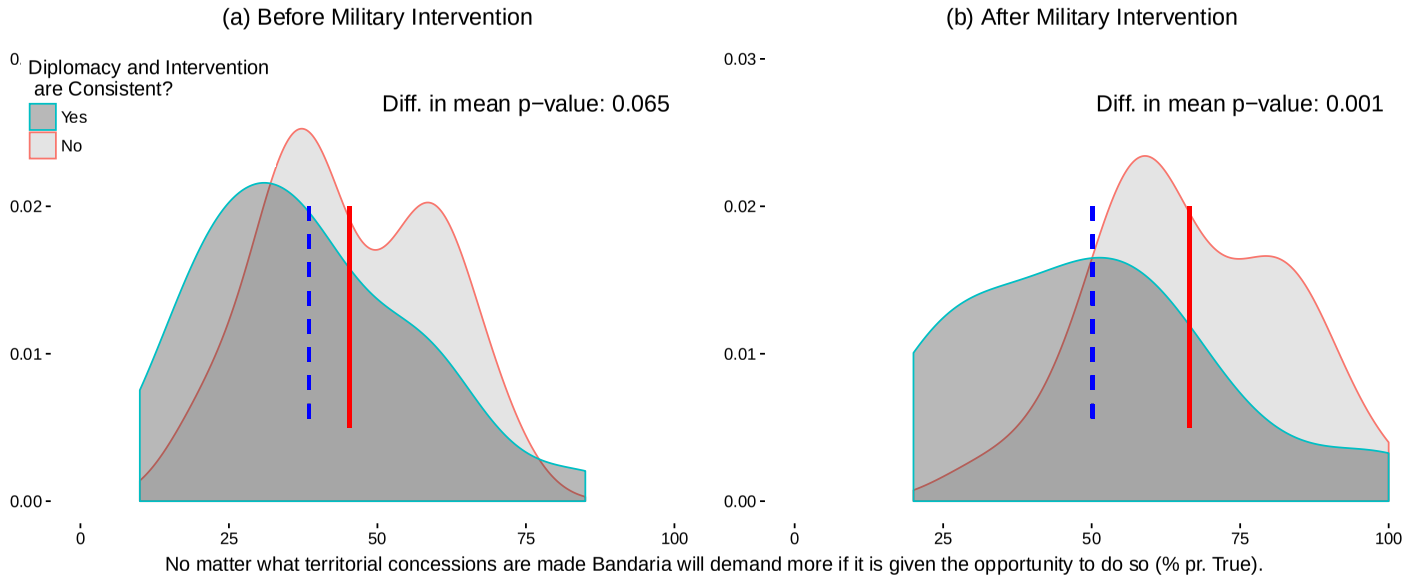


Figure 5: Does Inconsistent Behavior Alarm Subjects?

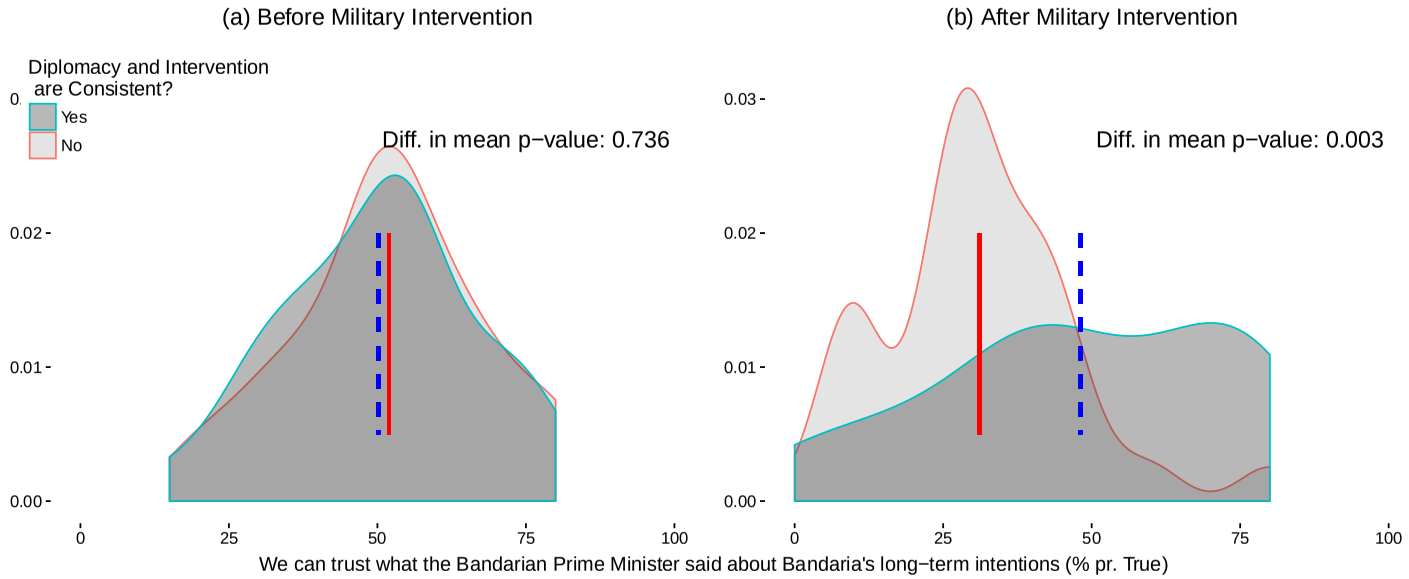


Figure 6: Does Inconsistent Behavior Lead to Less Trust?

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 6 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 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 7. 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.

3.5 Further Analysis

In Appendix E, 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, the controls are not significant and do not confound the treatment effect. An analysis of the covariance suggest that the treatment effect does not vary with the controls. 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. Taken together, these results demonstrate that the treatment effects subjects the same given different sampling methods and biographical backgrounds (elite and not), increasing our confidence that the result is not a function of the sampling method.

4 Conclusion

I argued that cheap, diplomatic reassurances are a critical piece of information during power transitions. At the onset of a power transition, rising powers promise their aims are limited by appealing to an underlying principle. They promise that once they achieve all the concessions related to that principle, they will peacefully integrate in the existing world order. Declining powers use these reassurances to evaluate the rising power's behavior from then on. They observe if military spending, demands and domestic policies are consistent

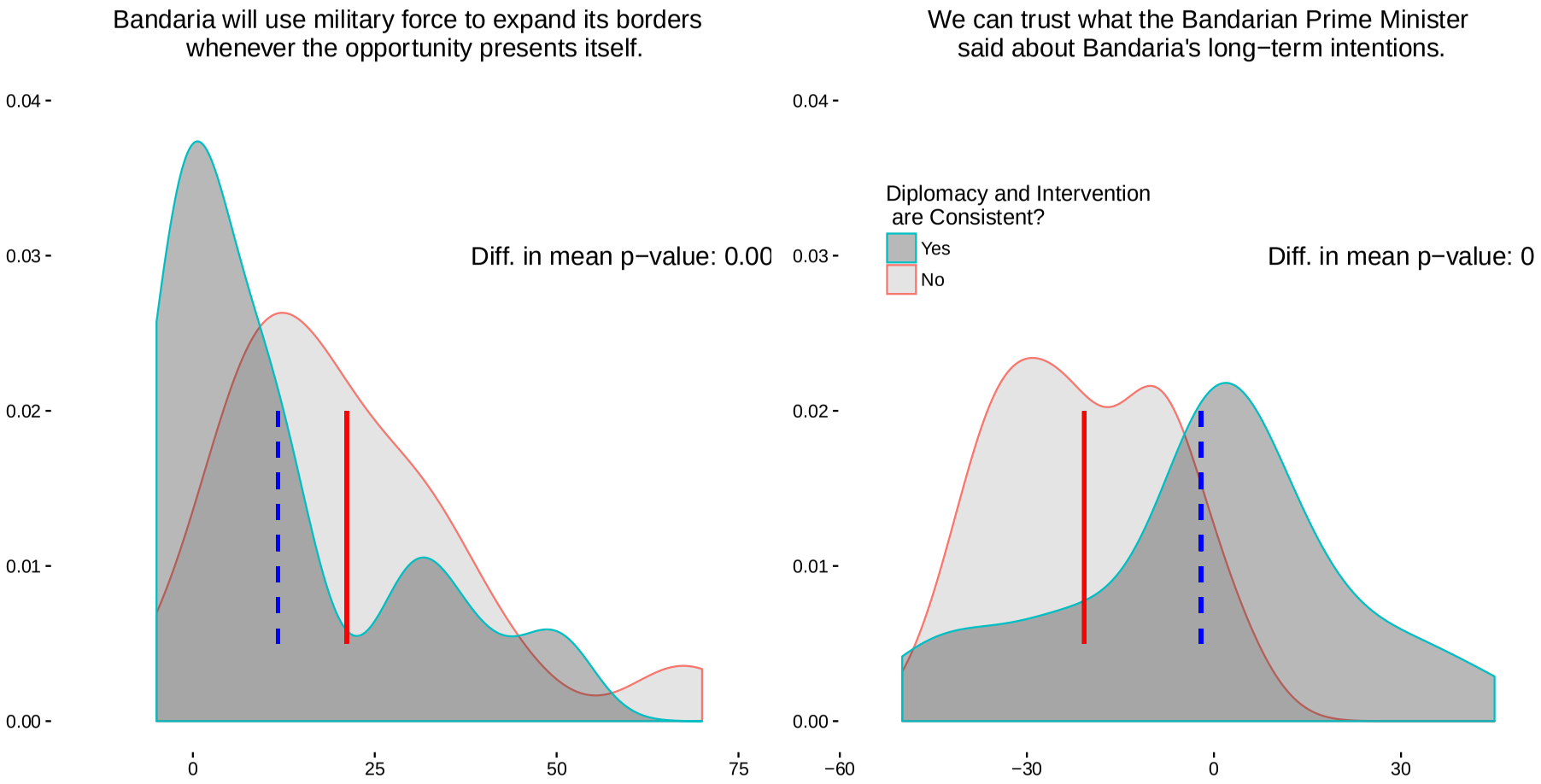


Figure 7: Does Inconsistent Behavior Lead Subjects to Change Beliefs Across Time?

with what was said. They trust rising powers when behavior is consistent, but become alarmed when behavior is not. I present unique experimental evidence that shows consistency between cheap diplomacy and action is precisely how foreign policy professionals make their assessments on average.

This finding implies that the critical period in Sino-American relations is approaching. The United States will soon discover if China's long-standing claims were genuine or if it was lying all along. Indeed, some have already determined that China is deceitful. This point in history will produce the highest probability of Sino-American crisis and we ought to prepare accordingly.

The logic of consistency between words and deeds may translate to other dynamic settings with uncertainty across multiple dimensions and shifting bargaining leverage. Such dynamics include how presidential candidates make campaign promises credible, or how the United States selects which terrorist organizations or nuclear aspirants to resist and which to ignore. Future research could explore these applications.

Finally, I pave the way for elite survey experiments to test formal models with uncertainty. Many observable behaviors (e.g. war) hinge on how policy-makers process information and form beliefs under uncertainty. Many theories make similar predictions about levels of uncertainty and conflict outcomes but rely on very different mechanisms. A good way to test which mechanism is at work is to observe how decision-makers actually evaluate evidence under these conditions.

References

Aldrich, R. J., R. Cormac, and M. S. Goodman (2014). *Spying on the World: The Declassified Documents of the Joint Intelligence Committee, 1936-2013*. London, UK.

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.

Battaglini, M. (2002, jul). Multiple Referrals and Multidimensional Cheap Talk. *Econometrica 70*(4), 1379–1401.

Broockman, D., J. Kalla, and P. Aronow (2015). Irregularities in LaCour (2014).

Brooks, S. (1999, oct). The Globalization of Production and the Changing Benefits of Conquest. *Journal of Conflict Resolution 43*(5), 646–670.

Carr, E. H. (1964). *The Twenty Years' Crisis, 1919-1939.* Harper & Row.

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.

Chakraborty, A. and R. Harbaugh (2007, jan). Comparative cheap talk. *Journal of Economic Theory 132*(1), 70–94.

Collett, J. L. and E. Childs (2011). Minding the gap: Meaning, affect, and the potential shortcomings of vignettes. *Social Science Research 40*, 513–522.

Cooper, R. W., D. V. DeJong, R. Forsythe, and T. W. Ross (1992, may). Communication in Coordination Games. *Quarterly Journal of Economics 107*(2), 739–771.

Crawford, V. and J. Sobel (1982). Strategic Information Transmission. *Econometrica: Journal of the Econometric Society 50*, 1431–1451.

Dafoe, A., B. Zhang, and D. Caughey (2016). Confounding in Survey Experiments.

Debs, A. and N. P. Monteiro (2014, jan). Known Unknowns: Power Shifts, Uncertainty, and War. *International Organization 68*(01), 1–31.

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. (1994, 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.

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*.

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.

Gleditsch, K. S. (2002). *All International Politics is Local: The Diffusion of Conflict*. Ann Arbor: University of Michigan Press.

Goemans, H. E. and K. A. Schultz (2013). The Politics of Territorial Disputes: A Geospatial Approach Applied to Africa.

Goodman, M. S. (2014). *The Official History of the Joint Intelligence Committee*. London, UK.

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.

Jackson, M. O. and M. Morelli (2011). The Reasons for Wars: An Updated Survey. *The Handbook on the Political Economy of War*, 34–57.

Jackson, M. O., H. Sonnenschein, and Y. Xing (2015, mar). The Efficiency of Bargaining with Many Items. *SSRN Electronic Journal*.

Jervis, R. (1978, jan). Cooperation Under the Security Dilemma. *World Politics 30*(02), 167–214.

Kaysen, C. (1990, apr). Is War Obsolete?: A Review Essay. *International Security 14*(4), 42–64.

Keohane, R. O. R. O. (2005). *After hegemony : cooperation and discord in the world political economy*. Princeton University Press.

Krainin, C. (2017, jan). Preventive War as a Result of Long-Term Shifts in Power. *Political Science Research and Methods 5*(01), 103–121.

Kugler, J. and D. Lemke (1996). *Parity and War: Evaluations and Extensions of the War Ledger*. University of Michigan Press.

Kurizaki, S. (2007, aug). Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy. *American Political Science Review 101*(03), 543–558.

Kydd, A. H. (2005). *Trust and Mistrust in International Relations*. Princeton, N.J.; Woodstock: Princeton University Press.

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*.

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.

Mitchell, S. M. and B. C. Prins (1999, mar). Beyond Territorial Contiguity: Issues at Stake in Democratic Militarized Interstate Disputes. *International Studies Quarterly 43*(1), 169–183.

Moravcsik, A. (1998, oct). Taking Preferences Seriously: A Liberal Theory of International Politics: Erratum. *International Organization 52*(4), 229.

Organski, A. F. K. and J. Kugler (1980). *The War Ledger*. Chicago: University of Chicago Press.

Powell, R. (1996, dec). Uncertainty, Shifting Power, and Appeasement. *The American Political Science Review 90*(4), 749–764.

Powell, R. (1999). *In the Shadow of Power: States and Strategies in International Politics*. Princeton, N.J: Princeton University Press.

Powell, R. (2006, jan). War as a Commitment Problem. *International Organization 60*(01).

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.

Reed, W. (2003, nov). Information, Power, and War. *American Political Science Review 97*(04), 633–641.

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.

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. (2014, jun). What's in a Claim? De Jure versus De Facto Borders in Interstate Territorial Disputes. *Journal of Conflict Resolution 58*(6), 1059–1084.

Schultz, K. and H. Goemans (2015). Aims, Claims, and the Bargaining Model.

Schweller, R. L. (1992, jun). Domestic Structure and Preventive War: Are Democracies More Pacific? *World Politics 44*(02), 235–269.

Snyder, J. (1993). *Myths of Empire: Domesitc Politics and International Ambition*. Ithaca, N.Y.: Cornell University Press.

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.

Tingley, D. H. and B. F. Walter (2011, aug). Can Cheap Talk Deter?: An Experimental Analysis. *Journal of Conflict Resolution 55*(6), 996–1020.

Trager, R. F. (2010, may). Diplomatic Calculus in Anarchy: How Communication Matters. *American Political Science Review 104*(02), 347–368.

Trager, R. F. (2011, jul). Multidimensional Diplomacy. *International Organization 65*(03), 469–506.

Trager, R. F. (2013, nov). How the scope of a demand conveys resolve. *International Theory 5*(03), 414–445.

Vasquez, J. A. (1993). *The War Puzzle*. Cambridge University Press.

Waltz, K. N. (1979). *Theory of international politics*. Reading, Mass: McGraw-Hill.

Woodward, E. L. (1970). *British foreign policy in the Second World War, Vol 4*. London, UK: H.M.S.O.

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

Yarhi-Milo, K. (2014). *Knowing the adversary : leaders, intelligence, and assessment of intentions in international relations*. Princeton University Press.

A Solution to Model

In this appendix I provide technical information for the game theoretic model provided in the paper. The appendix contains:

- A technical description of R's messaging strategy and O's offering strategy.
- A solution to the informative equilibrium defined in proposition 2.1
- A numeric example.

A word of advice: the numeric example is comprehensive. Most readers have told me that they understand the model quickly after reading it. It may be worth it to read the numeric example before reading through the model's solution.

A.1 A technical description of R's messaging strategy and O's offering strategy

I make precise R's messaging strategy and D's offering strategy across every period of the game. Let $y(\sigma_t()|T^*, V)$ be R's messaging strategy conditional on R's type and T_1^*. In this messaging strategy, all types of R mix evenly over all feasible messages in all but two periods of the game.[48] In two periods of the game, R sends a distinctive message. R always sends a distinctive message in the first period. In the first period, R sends a message $\sigma_1(\hat{V}_1)$ conditional on V. Then in the $f(\hat{V}_1)$th period, R sends a second message conditional on $V, \hat{V}_1$. To be clear, this second distinctive message arises in a period conditional on R's first period message $\sigma_1(\hat{V}_1)$.[49]

[48] I chose mixed off-path messages so that I could ignore messy equilibrium refinements. Instead, I can simply focus on R's messaging in the first period and the critical period. Similar to Chakraborty and Harbaugh (2007)'s result, an outcome-equivalent equilibrium exists where R sends an honest message in all rounds. To make the solution even simpler, I can assume that only the first period message is informative. But that would detract from an analysis of why R can't send a second message.

[49] Of main interest is the message R sends in period $t = T_1^*$. Yet if R values fewer than T_1^* issues, I give R the opportunity to send an alternative message given the period that she receives all the valuable issues that she reported in period 1.

R observes her type then if $f(V) \leq T_1^*$, R sends an honest message: $\sigma_1(\hat{V}_1) = \sigma_1(V)$ in the first period. R also sends an honest message in the period $t = f(\hat{V}_1)$ such that $\sigma_{f(\hat{V}_1)}(\hat{V}_1) = \sigma_{f(\hat{V}_1)}(V)$.

If $f(V) > T_1^*$, R sends a dishonest message $\sigma_1(\hat{V}) \neq \sigma_1(V)$ in the first period such that $f(\hat{V}) = T_1^* < f(V)$. The subscript on $\hat{V}_1$ makes clear that this was the type R signaled in the first period. Further, the signaled type $\hat{V}$ is a randomly selected sub-type of the true type such that $\hat{V} \subset V$. In period $t = f(\hat{V}) = T_1^*$, R sends a new message $\sigma_{T*}(\hat{\omega_{T*}})$ such that $f(\hat{\omega_{T*}}) = 4 * T_1^* \leq f(V)$. This message is the next largest super-type of the type that R originally signaled.

Let $O(y|\sigma_1(), h(Q_t))$ be D's offering strategy given R's first period message, and the history of offers. Once D observes R's first period message $\sigma_1(\hat{V}_1)$, D concedes issues $\hat{V}_1$ values 1 in random order (and with equal probability). Off the path, if D does not stop the game in period $t = T_1^*$, once D has conceded all issues that $\hat{V}_1$ values 1, D concedes all remaining issues in random order until D decides to stop the game.

A.2 Proposition 2.1: Effective Cheap-talk Equilibrium

Here I provide a technical solution for the equilibrium. Readers may want to skip ahead to a numeric example in section A.3 that lays out the model's logic quite clearly.

To start, I write out a technical version of equilibrium behavior that includes all off-path beliefs and a complete description of choices at every stage of the power transition.

Proposition A.1 *Suppose*

$$\frac{T_{\hat{V}_i}}{f(P) + c} \geq \lambda_{T1} \tag{12}$$

and

$$3T_{\hat{V}_1} > c \tag{13}$$

can be solved for any first period message $\sigma_1(\hat{V}_i)$ with $f(\hat{V}_i) = T_{\hat{V}_i}$. And suppose in period $t = T_{\hat{V}_i}$

$$\frac{3T_{\hat{V}_i}}{f(P) + c} \leq \lambda_{T*+1}, \tag{14}$$

then a critical threshold T_1^* emerges that implies the following equilibrium strategy always exists:

- *R observes her type, V, then sends a first-period messaging following $y(\sigma(V \leq T^*)|V)$:*
  - *If $f(V) \leq T_1^*$ R signals honestly $\sigma_1(\hat{V}) = \sigma(V)$.*
  - *If $f(V) > T_1^*$ R signals dishonestly $\sigma_1(\hat{V}) \neq \sigma_1(V)$ in the first round such that $f(\hat{V}) = T_1^* < f(V)$. Further, the signaled type $\hat{V}$ is a randomly selected sub-type of the true type such that $\hat{V} \subseteq V$.*
- *D processes the signal in one of two ways. If the signal implies: $f(\hat{V}) < T_1^*$, then*
  - *D adopts beliefs $\beta_t|\sigma_1(\hat{V}) \implies pr(V = \hat{V}) = 1$.*
  - *D offers R all the issues that R values in random order.*
  - *Off the path, if R does not stop the game following $f(\hat{V})$ concessions, D stops the game.*
- *If the signal implies: $f(\hat{V}) = T_1^*$, then*
  - *D adopts beliefs that R's true type is either $\hat{V}$ or some super-type of $\hat{V}$. D's beliefs are structured such that $pr(V = \hat{V}) > pr(V = \omega_j \supset \hat{V})$ for all super-types, ω_j.*
  - *D offers elements equal to 1 in $\hat{V}$ in random order.*
  - *R stops the game once D has conceded all the issues that R values if D has not stopped the game.*
  - *In period, $t = T_1^*$ D stops the game if R has not.*
- *Off the path, if R sends a first-period message that implies $f(\hat{V}) > T_1^*$, D stops the game in the first period.*

I assume that if D observes off-path behavior, D believes that R is the type that maximally profits from this deviation.

I show that proposition A.1 forms an equilibrium in three steps. First, I derive equilibrium conditions 12 and 14 that define D's incentives. Second, I show that neither player can profitably deviate when these conditions are satisfied and types of R that send a dishonest first period message prefers to invest and face war at T_1^* (equilibrium condition 13).

In equilibrium, if D observed a first-period message at the threshold: $f(\hat{V}) = T_1^*$, D's best reply was to set a stopping rule $r_D = T_1^*$. Condition 12 is based on D's first period preference for making T_1^* concessions under the assumption that in period T_1^* R will stop

the game with probability λ_{T1}, and that R will not stop the game with probability $1-\lambda_{T1}$. Further, if R does not stop that game in period T_1^* D selects competition:

$$\lambda_1(f(Q_1^D)-T_1^*)+(1-\lambda_1)(f(Q_1^D)-T_1^*-f(P)-c)>f(Q_1^D)-f(P)-c\equiv\frac{T_1^*}{f(P)+c}<\lambda_{T*1}, \tag{15}$$

Inequality 15 assumes that D prefers to compete in period $t=T_1^*$ if R does not stop the game. As a result, inequality 15 arises based on D's incentives in period T_1^* for competition. Define the type that is the next super-type of $\hat{V}$ as $\hat{V}_{+1}$. Given how types are constructed, $f(\hat{V}_{+1})=4*f(\hat{V}_1)$. Define the probability that R stops the game at the updated threshold λ_{T^*+1} D prefers to compete at T_1^* if R has not stopped the game if:

$$\lambda_{T^*+1}(Q_0-4T^*)+(1-\lambda_{T^*+1})(Q_0-4T^*-f(P)-c)\geq Q_0-T-f(P)-c\equiv\frac{3T^*}{f(P)+c}\geq\lambda_{T^*+1} \tag{16}$$

I'll now solve for D's beliefs λ_{T*1} in the first period and λ_{T^*+1} in the T_1^*th period, given R's equilibrium messaging strategy $y(\sigma_t()|T^*,V)$ and D's offering strategy $O(y|\sigma_1(),h(Q_t))$.

In equilibrium, R's strategy depends on the number of issues she values. There are two possible pathways depending on if $f(V)\leq T_1^*$, or $f(V)>T_1^*$. Types that satisfy $f(V)\leq T_1^*$ receive their maximum possible utility if they signal honestly. They receive every issue they value and face no cost of competition. Since they maximally profit on the path, they cannot profit from deviating.

Types $f(V)>T_1^*$ send a first period message $\sigma_1(\hat{V})$ such that $f(\hat{V})=T_1^*<f(V)$ and $\hat{V}\subset V$. I'll now show that no type $f(V)>T_1^*$ can profit from an alternative message give D's equilibrium reply. On the path, D processes R's first-period message $\sigma_1(V)$ and makes offers against issues that $\hat{V}$ values 1.

Suppose a type $f(V)>T_1^*$ sent a message that implied she valued more than T_1^* issues. She would face competition in the first period. This is clearly worse. Suppose she claimed to be a type that was not a subset of her true type ($\hat{V}\not\subseteq V$), then she would receive $\hat{V}$ worthless

concessions. At $t = f(\hat{V})$ D selects competition if R does not stop the game. Clearly, R does worse with this message. Suppose she sent a message that implied $f(\hat{V}) < T_1^*$. Then on the equilibrium path, she would receive fewer concessions before D shifted to competition. Clearly worse. There are no more deviations to consider.

I now analyze D's beliefs following R's first period message on the path. The purpose is to show that I can find a set of D's posterior beliefs λ_{T*1} for a specific threshold T_1^* that satisfies inequality 15. To do so, I'll conjecture that $T_1^* \in f\{\omega_i\}$ exists then show D's posterior beliefs satisfy given R's strategy on the path.

If D observes a first-period message below the threshold—$f(\hat{V}) < T_1^*$— D updates his beliefs such that $pr(V = \hat{V}|f(\hat{V}) < T_1^*) = 1$. The reason that D is completely persuaded by R's first-period message is that all types of R that value $f(V) \leq T_1^*$ send a unique, honest messages.[50] All types $f(V) \geq T_1^*$ pool with messages at T_1^*. It follows that $f(\hat{V}) < T_1^*$ must be honest.

Following a first period message $f(\hat{V}) = T_1$, D updates his beliefs to rule out all types other than $\hat{V}$ and all super-types of $\hat{V}$. D rules out types $f(V) \leq T_1$ except $\hat{V}$ because these types all send an honest message. D rules out types $f(V) > T_1$ except super-types of $\hat{V}$ because these types all send a message of a type nested within themself.

Further, D's belief that R signals honestly is:

$$\beta_1(V = \hat{V}|\sigma_1(\hat{V})) = \frac{P(\hat{V})}{P(V = \hat{V}|\Omega, y) + P(V \neq \hat{V}|\Omega, y)} \tag{17}$$

D's beliefs that R is dishonest, and has a true type that is the m-next largest super-type of $\hat{V}$ is:

$$\beta_1(V = \omega_{+m}|\sigma_1(\hat{V})) = \frac{P(\hat{V})}{4^m * (P(V = \hat{V}|\Omega, y) + P(V \neq \hat{V}|\Omega, y))} \tag{18}$$

This implies that following a message $\sigma_1(\hat{V})$ D is more confident that R is $V = \hat{V}$ than

[50] By unique, I mean that no type in the set $f(V) \leq T_1^*$ sends the same message as another type in that set.

any type $V \supset \hat{V}$. The reason is that only $V = \hat{V}$ sends an honest message with 100% probability. All super-types, mix over sub-types (which come in multiples of 4). It follows that:

$$\lambda_{T*}(|f(\hat{V}) = T_1^*) = \frac{4^{\log_4 N - \log_4 T_1^*}}{\sum_{m=0}^{\log_4 N - \log_4 T_1^*} 4^m} \tag{19}$$

Given that all types $f(\hat{V}) > T_1^*$ mix in this way,[51] we can solve for the smallest value $\lambda_{T1}|T_1^*, f(\hat{V}) = T_1^*$. For any T_1 and matrix size nxn, if I take the sequence to the limit of m and sum the probabilities, any $\lambda_{T1} \to 75\%$ (from above) as $nxn \to \infty$. This implies that for any T_1^*, D processes a first period message that implies $f(\hat{V}) = T_1^*$ such that $\lambda_{T*} > .75$. Plugging in $.75 = \lambda_{T1}$ into inequality 15, if $\frac{T_1}{f(P)+c} < .75$ can be solved for some $T_1 \in f(w\,\{\omega_i\})$, then inequality 15 can be satisfied. Notice that the smallest possible value for $T_1 = 1$. Thus, even with low competition parameters: $f(P) = 1$ and $c = 1/3$, inequality 15 can be satisfied for any matrix size nxn.[52]

Inequality 15 assumes that D will not stop the game for periods $t < T_1^*$.[53] However, at period T_1^* D will stop the game if R does not. For this to be true, it must also be true that inequality 16 is satisfied.

Suppose R plays a first period message $\hat{V}_1$ such that $f(\hat{V}) = T_1^*$. On the path, R stops the game at $t = T_1^*$ if $\hat{V}_1 = V$ and not otherwise.[54] Types $\hat{V}_1 \subset V$ do not stop the game and pool on a message $\sigma(\hat{V}_{+1})$ (the next largest super-type of $\hat{V}_1$). If R does not stop the game, D rules out the possibility that $V = \hat{V}$. But otherwise, D keeps the same ratio in his beliefs about whether R is the super-types of $\hat{V}$. D's posterior belief that R is $\sigma(\hat{V}_{+1})$ and will stop the game following $3T_1^*$ more concessions is:

[51] And types $f(V) = T_1^*$ send honest messages

[52] With the smallest matrix (4x4), I can satisfy inequality 15 with competition parameters: $f(P) = 1$ and $c = 1/4$. It is easier to satisfy as matrices grow.

[53] Clearly, D does not stop the game for periods $t < T_1^*$. On the path, all types that send a first period message $f(\hat{V}) = T_1^*$ play identical strategies until period $t = T_1^*$.

[54] If R stops the game, D's beliefs are irrelevant.

$$\lambda_{T*+1}(|f(\hat{V}) = T_1^*, r_R > T_1^*) = \frac{4^{\log_4 N - \log_4 T_1^* - 1}}{\sum_{m=0}^{\log_4 N - \log_4 T_1^* - 1} 4^m} \tag{20}$$

The denominator of λ_{T*+1} differs from λ_{T*} by 4^{-1} (so the denominator is smaller). The numerator includes one less sum because the top sum is reduced from $\log_4 N - \log_4 T_1^*$ to $\log_4 N - \log_4 T_1^* - 1$ (so the numerator is also smaller).

Overall, $\lambda_{T*+1} \geq \lambda_{T*}$. However, the difference is never that large. The most it can be is in the 4x4 matrix where $T_1^* = 4$. In that case, $\lambda_{T*} = 4/5$ and $\lambda_{T*+1} = 1$. But for any fixed $T_1^* < \frac{\text{nxn}}{4}$, as nxn $\to \infty$, $\lambda_{T*+1} - \lambda_{T*} \to 0$ (from above).[55]

Plugging in $\lambda_{T*+1} = 1$ into inequality 16, the outer bound of the inequality is $\frac{3T_1}{f(P)+c} > 1$.[56] But the largest possible $T_1 = \text{nxn}/4$. Subbing that in we get $\frac{3N}{4} > f(P) + c$.

Taking the outer bounds of inequalities 15 and 16 together, I can always solve them if $\frac{3N}{4} > f(P) + c > \frac{4}{3}$. To be clear, this implies that the inequalities are jointly solvable if D's cost of competition is less than 3/4 of D's total value of all of the territories, and slightly larger than any single issue. This is always true given the equilibrium conditions.

There are usually many values of T that might jointly satisfy these inequalities. But T_1^* is a specific value. Define T_1^* as the largest $T \in f(w\{\omega_i\})$ that satisfies inequality 15 and 16.

I'll now show that neither player can profit from deviating from the equilibrium described in proposition 2.1 given different stopping rules or offering strategies. First, I consider D's possible deviations. D can deviate by adjusting his stopping rule, or altering his offering strategy. By definition of inequalities 15 and 16 D cannot profit from setting an alternative stopping rule. So I focus on D's incentives to alter his offering strategy.

Suppose in some period, D deviated and offered R any issue that $\hat{V}$ valued 0. But R does not stop the game until R receives all her valuable issues. It follows that D cannot profit from this deviation. Either D keeps his stopping rule and faces competition. Or D delays stopping the game one period, and concedes one additional issue. Also, since all types of R

[55] When $T_1^* = \frac{\text{nxn}}{4}$ it implies that if R's first period message was dishonest, then R must be the greediest type because there are no other types.

[56] Notice that for values of $T_1^* < \frac{\text{nxn}}{4}$ that the condition is at least $\frac{3T_1}{f(P)+c} \geq 1$.

send mixed messages in every period until $t = \hat{V}$ and do not stop the game, then D can't learn any additional information from changing his offer before $t = \hat{V}$.

Turning to R's strategy. R can deviate by changing her stopping rule, or sending an alternative message. I've already argued that types $f(V) \leq T_1^*$ cannot profit from deviating because they receive their maximum payoff in equilibrium: $f(V)$. They receive all issues they value and face no costs.

Consider types that satisfy $f(V) > T_1^*$. On the path, they faces competition following T_1^* concessions and collect a total utility: $f(\hat{V} \times Q_{\omega 1}^R) + f(P \times Q_{\omega 1}^R) - c$. I'll show that no type can profit by deviating to a different stopping rule if equilibrium condition 13 is satisfied. At $t = T_1^*$, if $T_1^* + f(P) > f(V)$, then P only includes valuable issues and she collects: $T_1^* + f(P) - c$. In this case, at $t = T_1^*$, she ask for another concessions and face war if: $T_1^* + f(P) - c > T_1^* \equiv f(P) > c$, true by assumption. If $T_1^* + f(P) < f(V)$, then P must include some worthless issues and she receives: $f(V) - c$. In this case, at $t = T_1^*$, she ask for another concessions and face war if: $f(V) - c > T_1^*$. We know that $f(V) \geq 4 * T_1^*$. Thus, so long as $3T_1^* > c$ R cannot profit by stopping the game at T_1^* as stated in equilibrium.[57] In all periods $t < T_1^*$, no type $f(V) > T_1^*$ can profit from stopping the game because they receive valuable concessions if they don't.

Finally, I'll show that R cannot profit from sending an alternative message. Since R mixes messages in all but 2 distinctive periods, I focus on a message pair in periods $t = 1, f(\hat{V})$. No type $f(V) > T_1^*$ can profit from an alternative message that leads D to set a stopping rule at some period $t \leq T_1^*$.[58] On the path, D processes R's first-period message $\sigma_1(\hat{V})$ and makes offers against issues that $\hat{V}$ values. But $\hat{V} \subset V$. Thus R type-V is guaranteed T_1^* valuable concessions before D stops the game. Suppose R sent an alternative message that led D to set an earlier stopping rule, R must do worse because R gets less concessions. Suppose R

[57] Even when $3T_1^* > c$ does not hold, a very similar equilibrium emerges if $15T_1^* > c$. In it, all types pool their such that $f(\hat{V}) = T_1^*$. But at $t = T_1^*$, $\hat{V}$s next largest super-type, stops the game and pools with $\hat{V}$. All greedier types do not stop the game and face competition. As c grows, more greedier types pool at T_1^* and accept the status quo.

[58] After I discuss D's beliefs, I'll show that D cannot play a different message that induce more than T_1^* concessions.

sent an alternative message that led D to make T^* concessions of a type not nested in V. Then R would receive no valuable issues and still face competition at T_1^*.

Since T_1^* is the largest value that satisfies inequality 15, R cannot send a first period message that leads D to make more than T_1^* concessions by definition of T_1^*. Consider R sent a first period message $f(\hat{V}) > T_1^*$. D rules out the possibility that R is a type $f(V) \leq T_1^*$, because none of these types can profit from deviating. As a result, D knows that R values at least $3T_1^*$. Clearly, no set of plausible posterior off-path beliefs can induce D to make concessions because T_1^* is the largest value that satisfies inequality 15 and 16.[59]

For the same reason, R also cannot induce more concessions from sending some first period message $f(\hat{V}) < T_1^*$ followed by a $\hat{V}_2$th period message that leads to more concessions. Again, only types $f(\hat{V}) > T_1^*$ can profit from deviating and so following any off-path message, D believes that R is one of these types. But by definition of T_1^*, D is unwilling to make concessions to these types if they expose themselves in period T_1^*. Clearly, D is unwilling to do it in an earlier period.

A.3 Numeric Example

To make the mechanism concrete, I'll illustrate the results through a numeric example.

In this example, R and D bargain over the 8x8 matrix (64 total issues) with competition parameters $P = 8, c = 5$. Based on the distribution of types there are 85 possible value matricies (ω_i). These types are summarized in Table 2.

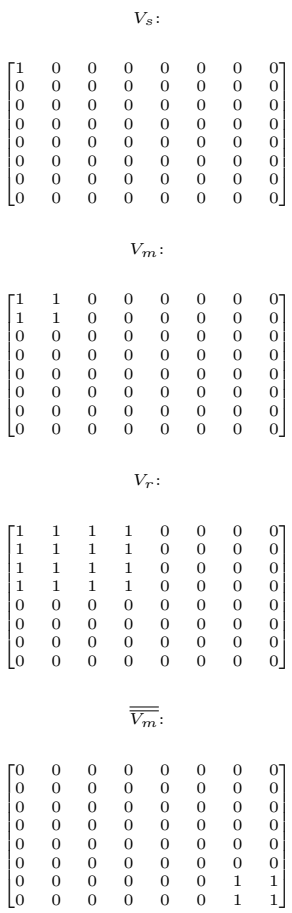
Table 2: Type-space in 8x8 numeric example

	Global hegemon	Regional hegemon	Multi-issue	Single-issue
No. issues valued	64/64	16/64	4/64	1/64
No. of types in $\Omega(w)$	1/85	4/85	16/85	64/85

The matrix size is 8x8 (64 total issues). In the example I assume $c = 3$, $f(P) = 10$.

[59] For example, D1 or D2 refinements would drive D to competition.

Figure 8: Different types of R in the numeric example



Notice that $V_s \subset V_m \subset V_r$. However $\overline{\overline{V_m}}$ values different objectives.

Throughout the example, I refer to four possible realizations of R's type, which I depict in Figure 8. Let V_s be the single-issue type that only values the element (1,1). Let V_m be the multi-issue type that values the 4 issues in the top left quarter of Q (that is, $f(V) = 4$). Let V_r be the regional hegemon that values the top left quarter of Q (that is, $f(V) = 16$). Finally, let $\overline{\overline{V_m}}$ be the multi-issue type that values the bottom right four issues.

For now, let's conjecture that T_1^* exists and equals 4. That is, D is willing to offer R four concessions then switch to competition if R does not stop the game, given that R sends a first period message $f(\hat{V}) = 4$.

If $T_1^* = 4$, then the game can take three different pathways depending on if R values more,

less or exactly 4 issues. I'll use the different types to illustrate what happens depending on if states want more, less or exactly this threshold.

- If Nature draws V_s (values less than $T_1^* = 4$) then in the first period V_s sends an honest first-period message, and D updates his beliefs such that $pr(V = V_s|\sigma_1(V_s)) = 1$. D offers element $(1, 1)$ that V_s values. In the second period, V_s stops the game and accepts the status quo.

- If Nature draws V_m (values exactly $T_1^* = 4$) then in the first period V_m sends an honest first-period message, and D updates his beliefs such that $pr(V = V_m|\sigma_1(V_m)) = 16/21$ and $pr(f(V) > f(V_m)|\sigma_1(V_m)) = 5/21$. In the first four periods, D offers the four elements that V_m values in random order. In the fifth period, V_m stops the game and accepts the status quo. The game passes peacefully.

- If Nature draws V_r (values more than $T_1^* = 4$) then in the first period V_r sends a dishonest first-period message $\hat{\omega_1} =\subset V_r$ such that $f(\hat{\omega_1}) = T_1^*$, and D updates his beliefs such that $pr(V = \hat{\omega_1}|\sigma_1(V_m)) = 16/21$ and $pr(f(V) > f(\hat{\omega_1})|\sigma_1(\hat{\omega_1})) = 5/21$. In the first four periods, D offers the four elements that V_m values in random order. In the fifth period, V_r does not stop the game. D's updated beliefs are such that D prefers competition to making additional offers. The game ends in fifth period competition.

For this set of behaviors to hold up, two surprising things must be true. First, D must be so confident that R's first-period message is honest that D is willing to concede everything that R claims to value in the first period, rather than compete. The conventional wisdom is that this should not happen because R faces strong incentives to mis-represent and D should expect the worst. However, I'll show that R's first-period message is so effective that D updates his beliefs enough to make offers in the first period. D prefers to make concessions for the first T_1^* periods if:

$$(64 - T_1^*)(1 - \lambda_{T1}) + (64 - T_1^* - 5 - 8)(\lambda_{T1}) > (64 - 5 - 8)\lambda_{T1} > \frac{T_1^*}{13} \tag{21}$$

The left hand-side is D's benefit from making T_1^* concessions. The first term, is what D gets if R stops the game following $T_1^* = 4$ concessions, multiplied by D's expectation that R's first period signal is honest $(1 - \lambda_{T1})$. The second term, is what D gets if R does not stops the game following T_1^* concessions and D chooses competition at $t = T_1^* + 1$, multiplied by D's expectation that R's first period signal is dishonest (λ_{T1}).[60] The right hand side is D's benefit from first-period competition.

Second, D must prefer competition once R reveals that her first-period message was dishonest in period $t = T_1^*$. In period $t = T_1^*$, D prefers competition if:

$$(64 - 3T_1^*)(1 - \lambda_{T*+1}) + (64 - 3T_1^* - 5 - 8)\lambda_{T*+1} < (64 - T_1^* - 5 - 8) \tag{22}$$

$$\frac{3T_1^*}{13} > \lambda_{T*+1} \tag{23}$$

Here the subscript on λ_{T*+1} implies it is the message R sends in period $T_1^* + 1$. For this inequality to hold, it must be that R cannot convince D that her aims are limited in the T^*th period. It is surprising that R cannot send a second credible message given that R's first period message was so effective. After all, if R's first period message convinced D that R had limited aims, then why can't R update that message at $t = T_1^*$, and convince D that she wants just a little bit more?

I now illustrate why inequalities 21 and 23 jointly hold To start, I'll consider how different type of R wants to signal given D's equilibrium response. Any type that values $f(V) \leq 4$ receives her maximum possible pay-off if she signals honestly. If V_s signals honestly, D concedes element $(1, 1)$ in the first period, and V_s pays no competition costs. Thus, V_s cannot profit from sending an alternative first-period message. V_m also receives her maximum possible pay-off if she signals honestly. The reason is that V_m values four issues but D will not stop the game until the fifth period.

[60] Of course, this inequality assumes that D knows he can concede R exactly what R values.

There are other messages these types could send to avoid first-period competition. However, these alternative messages lead to worse outcomes. V_s could send an off-path message $\sigma_1(\overline{\overline{V_m}})$. D would process this message and offer the four issues in the bottom right corner of Q. But V_s does not value these issues. Thus, this message would produce 4 worthless concessions for V_s.

This counter-factual message highlights how cheap-talk can overcome the informational challenges that complex preferences create. In the babbling equilibrium, D could not figure out which concessions R valued. Even if R valued no more than 4 issues, D chose competition in the first period because D was unlikely to correctly guess which issues R cared about. Both R and D did worse from failing to coordinate over R's preference order. In the cheap-talk equilibrium, R wants to reveal which issues she values to guarantee valuable concessions and D wants to make these concessions.

Types that value more than four issues face different incentives. For example, V_r wants to avoid competition for as long as possible to maximize the number of concessions that she receives. But if V_r sent an honest first-period message she would face competition. Thus, V_r must under-state her aims if she is to avoid competition in the first period. The best message that V_r can send is any message from the four multi-issue types nested within herself. This best message implies that V_r does worse by sending any other message, but also that V_r is indifferent between sending any of these 4 messages (and therefore can mix over them).

To see that V_r does worse by sending any other message, consider two counter-factual messages. V_r could signal that she was $\sigma(V_s)$. However, this would produce only 1 concession before competition. Clearly, V_r does better by signaling a multi-issue type. But not all multi-issue messages produce the same amount of utility for V_r. If V_r sent a first period message $\sigma_1(\overline{\overline{V_m}})$ she would receive 4 worthless concessions. At $t = 5$, she is no better position than she was at the beginning of the game.

V_r is indifferent between sending any of the multi-issue types nested within herself because each of these messages guarantees V_r exactly four valuable concessions.

D can exploit these different incentives to make nuanced inferences from R's first period message. I'll illustrate the intuition by studying what happens when Nature draws V_r, and V_r sends an equilibrium message $\sigma_1(V_m)$. First, I'll show that D is optimistic that R values 4 issues following a dishonest first period message $\sigma(V_m)$ (recall $f(V_m) = T_1^* = 4 < f(V_r)$). Second, I'll show that in the fifth period, R cannot send another message that induces D to keep cooperating. As a result, D fights in the fifth period if R does not stop investing.

Suppose D observes a first period message $\sigma(V_m)$. He can rule out the possibility that R is a single-issue type because these types would have signaled honestly. He can also rule out all types whose interests do not intersect with V_m (e.g., $\overline{\overline{V_m}}$). These types would have send a multi-issue message. However, they would have sent an honest message that guaranteed them valuable concessions. As a result, D infers that R is one of three types: V_m, V_r and the greediest type that values every issue (V_G). Notice these three types are nested within each other.

Before R sent a message, D thought it was equally likely that R was V_m, V_r or V_G. D's prior beliefs were $pr(V = V_m) = pr(V = V_r) = pr(V = V_G) = 1/85$. Following R's first period message $\sigma_1(V_m)$, D's posterior beliefs are re-weighted in favor of V_m such that $pr(V = V_m) = 4 * pr(V = V_r) = 16 * pr(V = V_G)$.

D updates his beliefs because V_m, V_r, V_G are not equally likely to send a message $\sigma(V_m)$. Recall that V_r wanted to signal a multi-issue type nested within herself. But V_r was indifferent between sending a message $\sigma(V_m)$, and a message that implied V_r was another multi-issue type nested within herself (for example, the type that valued issues $(3,1),(3,2),(4,1),(4,2)$). In equilibrium, V_r chooses from the four multi-issue types nested within herself at random, each with one quarter probability. In contrast, V_m must send an honest message to receive valuable concessions.[61]

By Bayes' Rule, D's posterior belief that R is honest is:

[61] I showed above that an honest message uniquely maximized her utility.

$$pr(V = V_m|\sigma_1(V_m)) = \frac{pr(V = V_m) * pr(\sigma_1(V_m)|V = V_m)}{pr(V = V_m) * pr(\sigma_1(V_m)|V = V_m) + pr(V = V_r) * pr(\sigma_1(V_m)|V = V_r) + pr(V = V_G) * pr(\sigma_1(V_m)|V = V_G)} \quad (24)$$

$$\frac{1/85 * 1}{1/85 * 1/16 + 1/85 * 1/4 + 1/85 * 1} = 16/21 \quad (25)$$

The denominator is D's prior belief that R is type V_m (1/85) multiplied by the probability that R is type V_m and V_m sent a message $\sigma_1(V_m)$. The numerator is made up of three terms: the probability that R is the greediest type multiplied by the probability that R is the greediest type and that type sent the message $\sigma_1(V_m)$; the probability that R is V_r, multiplied by the probability that R is V_r and V_r sent the message $\sigma_1(V_m)$; and the probability that R is V_m and V_m sent an honest message.

D's posterior belief that R is V_r given a message V_m is $pr(V = V_r|\sigma_1(V_m)) = \frac{1/85*\mathbf{1/4}}{1/85*1/16+1/85*1/4+1/85*1} = 4/21$. D's posterior belief that R is V_G given a message V_m as $pr(V = V_G|\sigma_1(V_m)) = \frac{1/85*\mathbf{1/16}}{1/85*1/16+1/85*1/4+1/85*1} = 1/21$.

D re-weights his beliefs because $\Omega(w)$ contained both variation in scope and preference order. Types that valued only a few concessions face incentives to identify the few concessions they value. In this example, V_m only valued four issues. She could only receive them if she sent an honest message. In contrast, V_r can send 4 different messages that will guarantee her 4 valuable concessions. V_r's incentives are to receive any four valuable concessions and hide her true intentions for as long as possible.

This implies that $pr(V = V_m) = \lambda_{T1} = 16/21$. Plugging this value into inequality 21, I can satisfy it if $T_1^* = 4$ as desired.

On the path, V_g, V_r, V_m play identical strategies for four periods. So there is nothing that D can learn until period T_1^*. At that point, only V_m stops the game and the others reveal that their initial message was dishonest.

In deriving inequality 21, I assumed that D would stop the game if he discovered at T_1^* that R's initial message was dishonest. This implied inequality 23 must also be satisfied. To

satisfy this inequality, R was unable to convince D that she wanted just a little bit more in period T_1^*.

One might wonder if R's first round message is so effective, then why can't R just send another effective message at $t = T_1^*$? In our example, why can't V_r send a new message that persuades D she is the next largest type? Two factors work against V_r sending a second effective message in the 5th period. First, since types are nested within each other the next largest type always wants more than the type that came before it. In the first period, D only needed to make 4 concessions before he discovered if R was honest. But in the fifth period, D anticipates that the next largest type wants 12 more concessions.

But this escalating threshold cannot completely explain why competition emerges. In the numeric example, D's cost of competition is $c + f(P) = 13$. If R sent a credible message $\sigma(V_r)$ in the fifth period, then D would prefer to make 12 more concessions rather than select competition at a cost of 13.[62] Thus, the question remains: why can't R send a second message that induces continued cooperation?

R's first period message altered D's beliefs because only the honest type sent the observed message with certainty. Greedier types chose their message at random from their sub-types. But once R sent this first message, D ruled out all types whose interests did not intersect with $\hat{\omega}$. In our example, once R sent the first message $\sigma_1(V_m)$, D ruled out all types that were not V_m, or super-types of V_m (i.e V_r, V_G). Thus, R's first period message eliminates all variation in preference order and all remaining types send identical messages. In the fifth period V_G and V_r send message $m(V_r)$ with 100% probability. As a result, once R invests in the fifth period, D rules out the possibility that R is V_m but keeps the ratio of beliefs $pr(V = V_r) = 4 * pr(V = V_G) \implies \lambda_5 = 1/5$. This is not enough for D to make additional concessions. As a result, inequality 23 is satisfied and D prefers competition to making additional offers at T_1^*.

[62] In fact, if R could send a fifth period message that was just as effective as R's first period message D would make additional offers.

B Model Robustness

Available on market materials tab of my website.

C Survey Text

Below is the text that subjects observed 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.

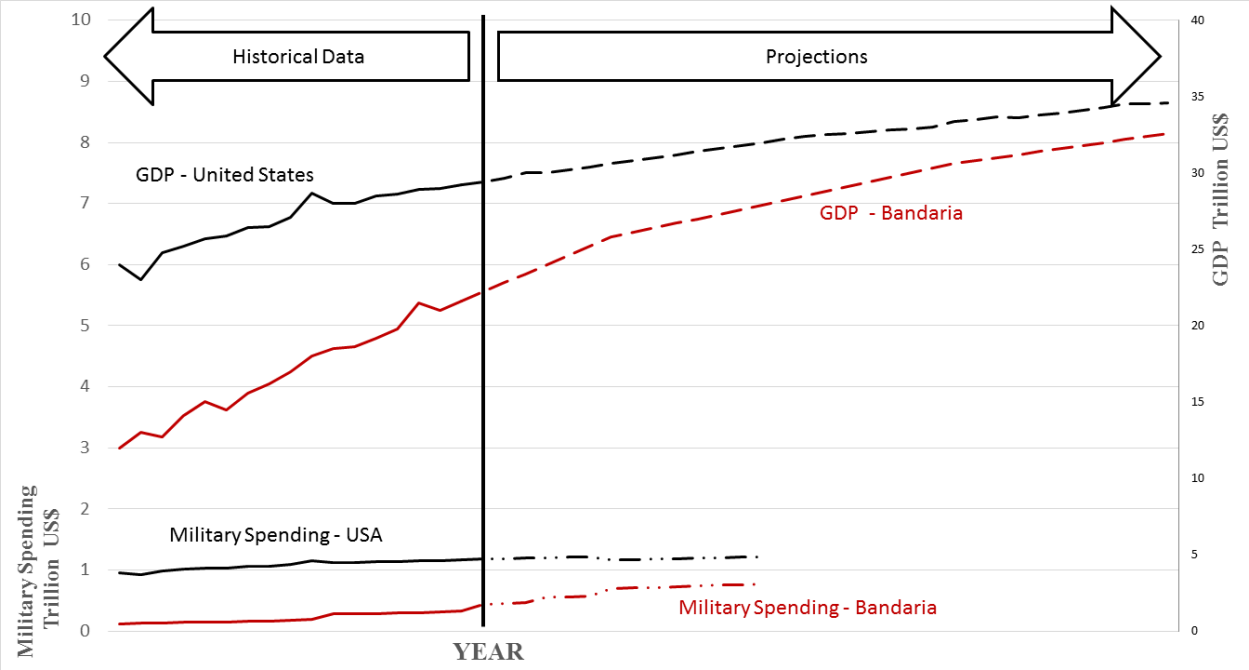
Evaluate the following information and provide an assessment of Bandaria's long-term intentions.

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

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 Experimental design

I provide additional information about design choices I made in the experiment. Appendix C presents the final text of the vignette that subjects received.

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.[63] 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. In the snowball sample, I asked my professional network to extend a solicitation email to subjects that they believed fit my sampling frame.

Table 3 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

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

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 3: Sampling Method

	Response to Strategic Intentions	Response to Resolve
	(1)	(2)
Fought for Core Interests	10.426**	−16.880**
	(3.759)	(5.100)
Institutional Sample (Y/N)	−2.354	−3.465
	(3.834)	(5.202)
Constant	78.236**	68.128**
	(3.171)	(4.303)
Observations	60	60
Note:		*p<0.05; **p<0.01

Figure 11(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 11(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.

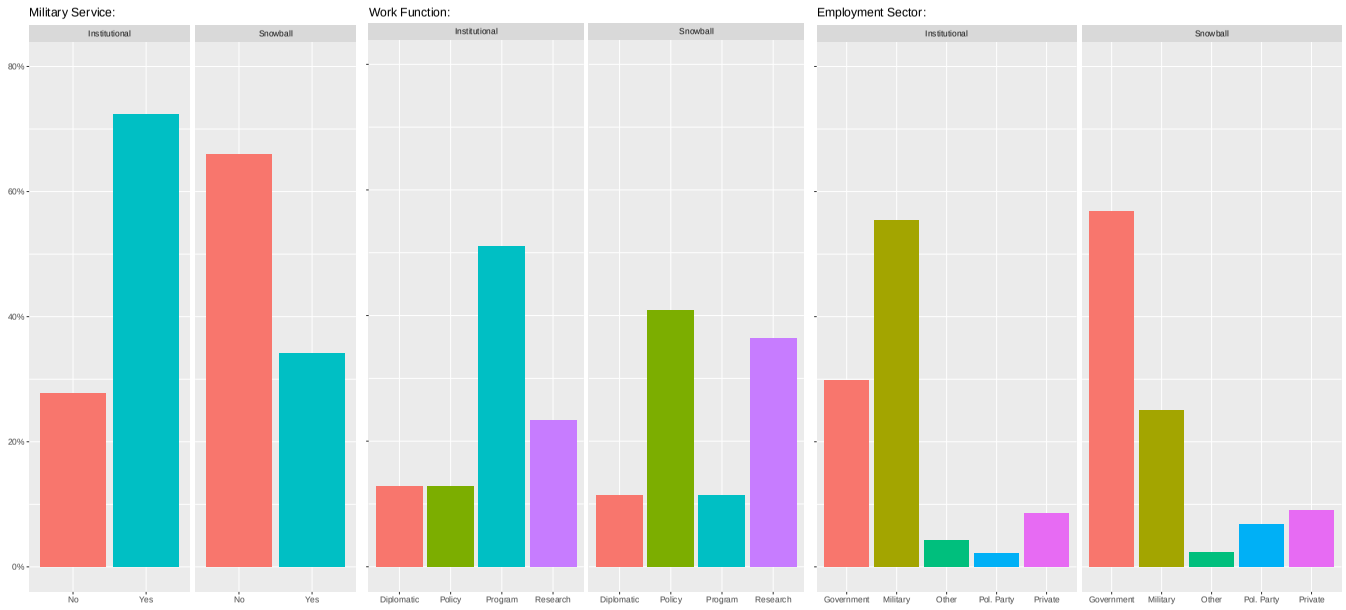
D.1.2 Solicitation Email

The following text is the solicitation email sent to subjects from the snowball sample:[64]

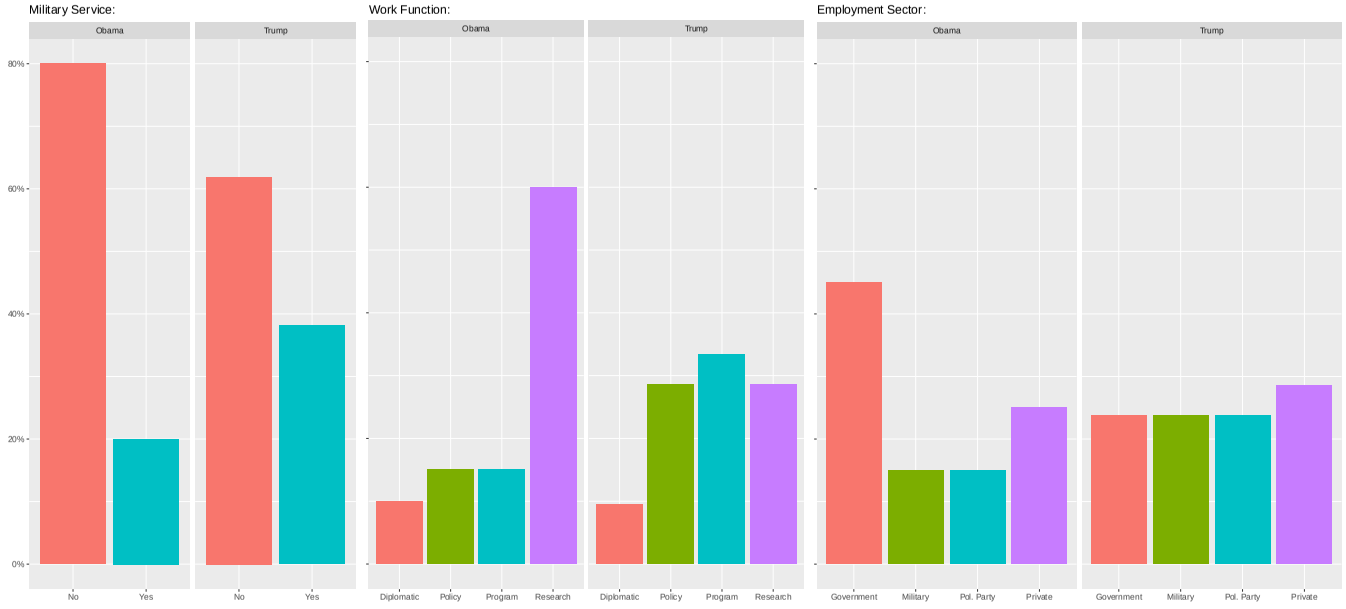
Please take part in a study that simulates a foreign policy assessment. We need foreign policy,

[64]The Solicitation for the institutions is similar, and therefore omitted.

Figure 11: 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.

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]

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 4 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 4: 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 posible 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,[65] and completed the survey,

[65] i.e subjects I am confident are elites.

32 failed at least one of the attention checks. Table 5 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 5: Elites Who Passed Attention Checks?

D.4.2 Meta-data

Table 6 reports the time it took subjects to respond and the length of their text responses.

Table 6: 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

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.[66] 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.[67] 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.

[66]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.

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

Table 7: 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 8: 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 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:

$$Trust_i = \beta_0 + \beta_1 Consistent + \beta_k Control_j + \epsilon \tag{26}$$

$$Opportunistic_i = \beta_0 + \beta_1 Consistent + \beta_k Control_j + \epsilon \tag{27}$$

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 9: Sampling Method

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

Table 10: Work Function

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

Table 11: Employment Sector

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

Table 12: Military Service

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

Table 13: American Citizenship

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

Table 14: Highest Ranking Official

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

Table 15: Time to Complete Survey

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

E Further Experimental Analyses

I now report additional experimental results in support of my argument. In an online Appendix available on my website I also report evidence from a post-survey questionnaire designed to rule out several alternative explanations.

E.1 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:

$$Trust_i = \beta_0 + \beta_1 Consistent + \beta_k Control_j + \epsilon \tag{28}$$

$$Opportunistic_i = \beta_0 + \beta_1 Consistent + \beta_k Control_j + \epsilon \tag{29}$$

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 16: Sampling Method

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

Table 17: Work Function

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

Table 18: Employment Sector

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

Table 19: Military Service

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

Table 20: American Citizenship

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

Table 21: Highest Ranking Official

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

Table 22: Time to Complete Survey

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

E.2 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:

$$sS \cap eE > cS \cap cE. \tag{30}$$

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.

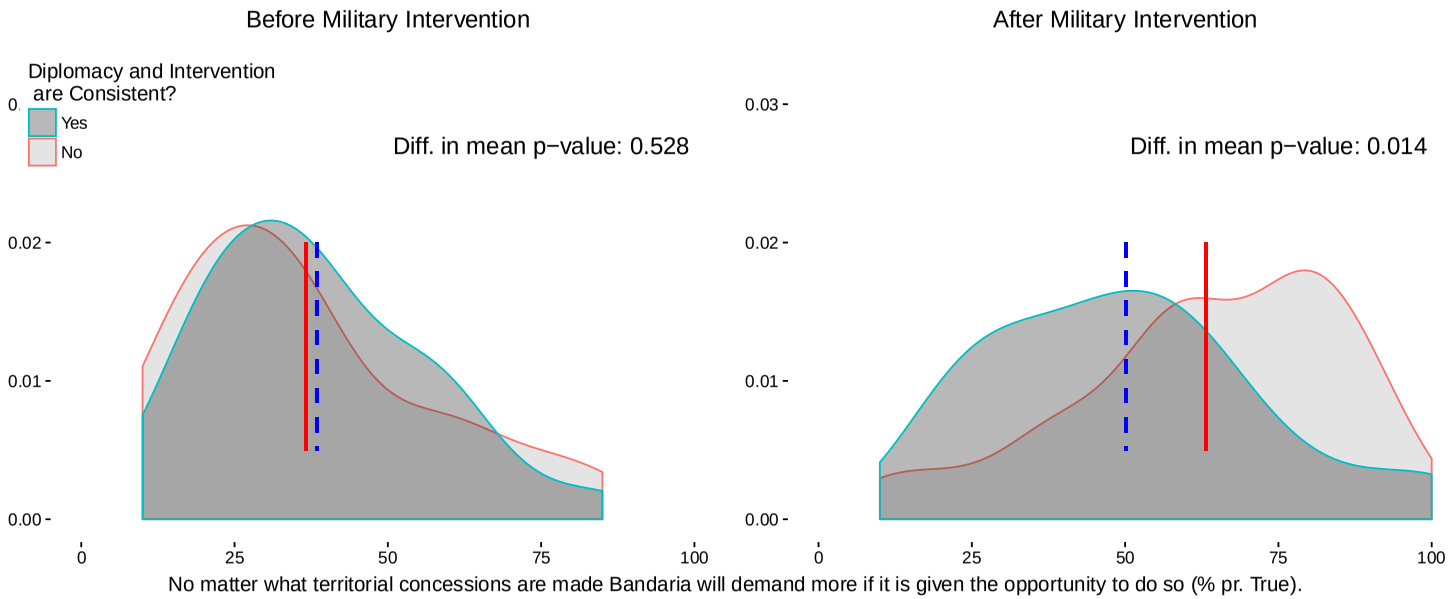


Figure 12: Does Cheap Diplomacy Influence Beliefs About Limited Aims?

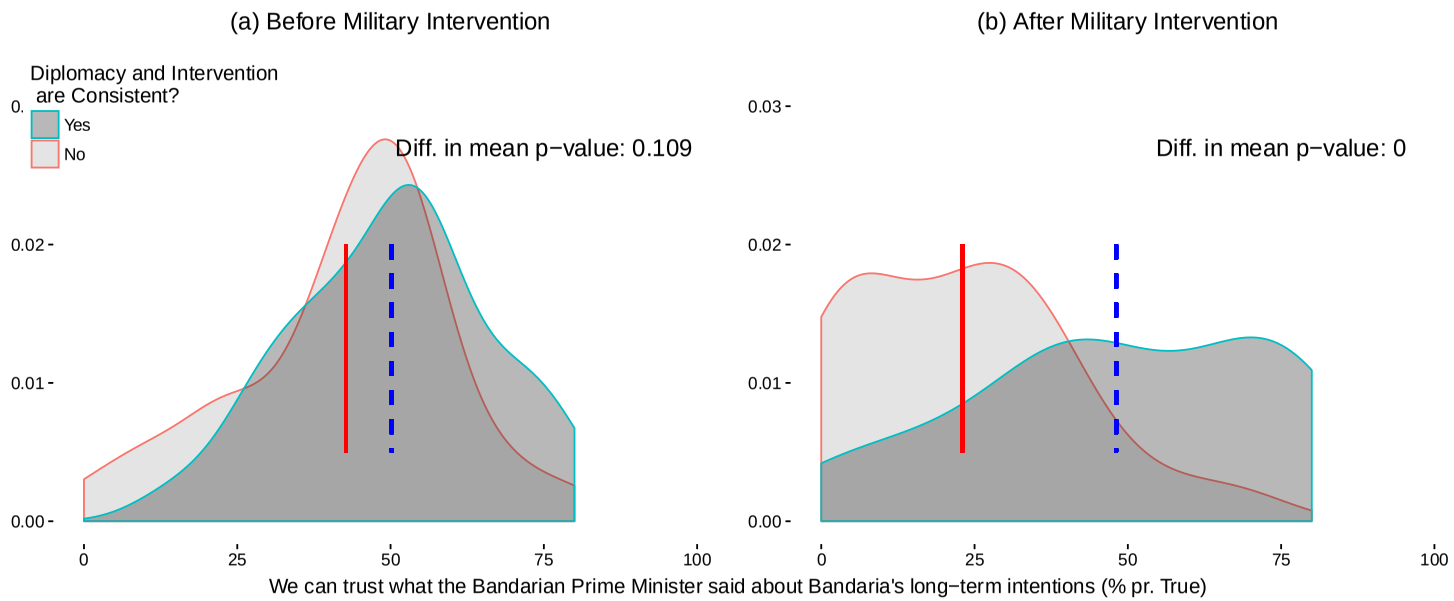


Figure 13: Does Inconsistent Behavior Lead to Less Trust?

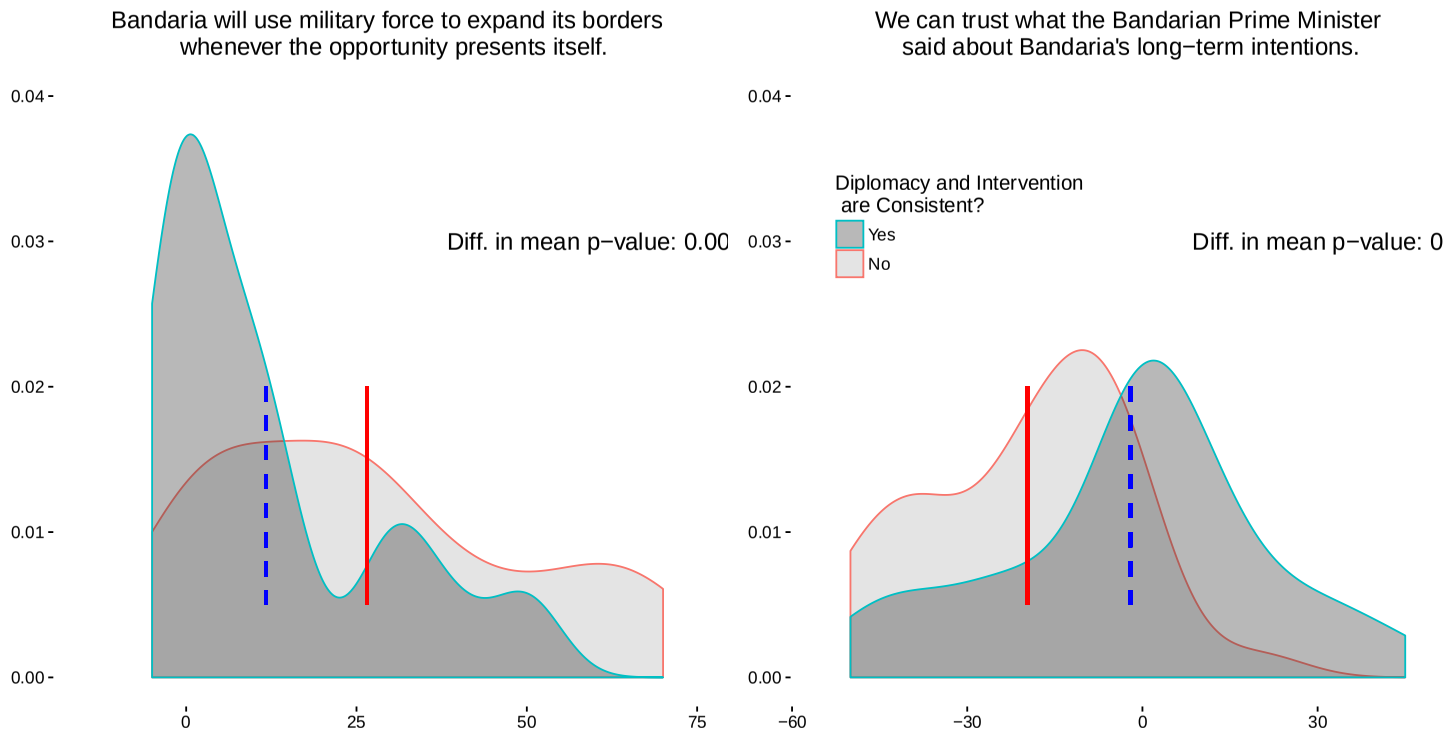


Figure 14: Does Inconsistent Behavior Lead Subjects to Change Beliefs Across Time?