Explaining cooperation, credible costless communication, coordination and delayed competition under the reassurance dilemma

Word Count: 12,996

September 2, 2019

Abstract

How do challengers to the status quo reassure great powers that their long-term motives are limited in repeated interactions? I offer a formal theory of reassurance in long-term interactions with repeated demands to explain: (1) how challengers exploit costless messages early on to re-assure great powers their intentions are limited even as they rapidly militarize and make violent demands; and (2) why great powers infer a challenger is aggressive following specific crisis episodes, and eventually turn to competition. I argue that states are motivated by different principles (e.g. nationalism, revenge, security) and each principle implies that the challenger care about specific territories. Great powers do not know what principle motivates the challenger, but they do know that there is only a finite number of plausible principles, and also which specific territories correspond to which principle. For example, British elites did not know if Hitler was motivated by nationalism or not. But they knew where all the Germans lived in Europe and exploited that information to evaluate Hitler's motives. The limited set of motives creates discontinuities in the number of issues that the challenger could want, and this provides nuanced opportunities for signaling across time. I integrate this theory of motives into a model of reassurance to make predictions about when great powers shift from cooperation to competition based on qualitative differences in the demands that challengers make and not the scope of their demands or rates of militarization. I validate my predictions and assumptions with archival research on Anglo-Soviet bargaining.

Introduction

How can great powers evaluate the long-term, strategic intentions of challengers they face? This question dictates patterns of cooperation and competition during power transitions (Yoder, 2019), arms races (Debs & Monteiro, 2014), enduring rivalries (Kydd, 2005), security dilemmas (Jervis, 1978) and broader interactions in world politics (Glaser, 2010; Mearsheimer, 2001).[1] It is also policy relevant: American foreign policy-makers realize that inferring "Chinese intentions is the single most *difficult* and *important* task we face."[2] The task is *important* because US foreign policy depends on it. The US must choose between incremental concessions over time, or costly competition designed to deny China the opportunity to make future demands. The task is *difficult* because Chinese elites faces a reassurance dilemma: They claim they want a peaceful rise, and they may truly hold limited aims. But American policy-makers are uncertain if this claim is genuine, or if Chinese elites are understating their motives to avoid competition.

For a long time we thought that challengers faced insurmountable incentives to understate their motives (Mearsheimer, 2001; Carr, 1964). But through history, states with limited conflicting preferences came to trust each other. For example, the British avoided competition with the United States circa 1900 because they realized that the United States held limited aims. Recently, researchers have discovered that challengers can underarm (Kydd, 2005; Axelrod, 1980), or forgo making demands (Yoder, 2019) to credibly communicate limited motives. These theories of costly re-assurance explain how trust forms in short (usually two-period) interactions.[3] However, they cannot explain why great powers remain calm as they observe challengers make repeated territorial demands, and rapidly militarize. For example, in 1932 British elites were uncertain if Hitler was a "madman" bent on world domination, or if he sought limited objectives. They worried when Hitler annexed the Rhineland, parts of

[1] This question is not relevant in crisis bargaining (Fearon, 1995), when states care about overstating their resolve to fight for a specific issue (Kertzer, 2016), rather than under-stating their interest in taking many issues.

[2] Quote from interview with former CIA Director for Analysis, Mark Lowenthal.

[3] Although Yoder (2019) describes a power transition, his model includes only two periods.

Austria and Czechoslovakia. But Hitler allayed their fears through diplomacy. During private exchanges he convinced three British prime ministers that once he satisfied Germany's nationalist goals he would peacefully integrate into European society. Only once Hitler violated the Munich Agreement did British elites 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."[4] With Hitler fresh in their minds, British and American leaders believed that Stalin's intentions were limited up until the Iran Crisis (1946).

Theories of rational reassurance cannot explain two features of these cases.[5] First, theories of costly re-assurance argue that great powers draw stronger inferences when the challenger makes large, or strategically important demands, or rapidly militarizes. But through history *how much* a challenger asked for mattered less than *what* they asked for. British elites inferred that Hitler was greedy because they believed his demands in Czechoslovakia and Poland were different from his demands in Austria, and therefore a more informative signals of his aggressive intentions. More general evidence of repeated great power interactions suggests that some issues matter more than others (Schultz, 2017). But across cases, there is no obvious pattern about which violent events great powers choose to draw inferences from and which they ignore (Press, 2007; Hopf, 1994).

Second, scholars are suprised by effective diplomacy (Yarhi-Milo, 2014). Hitler and Stalin used costless diplomatic messages to reassure British elites that their motives were limited, even as they make repeated territorial demands and rapidly militarize (Holmes, 2013; Yarhi-Milo, 2014). Yet unlike crisis bargaining interactions where cheap-talk can be effective (Trager, 2013; Kurizaki, 2007; Sartori, 2005), challengers in reassurance games understate, not overstate, their intentions. As Hitler did, even the most aggressive challengers argue that militarization is necessary to take a few valuable concessions but promise that their in-

[4]Nicholson Diaries, 27 Feb 1945. Discussion about Soviet strategic intentions.
[5]And many other cases. See Kennedy (1989); Copeland (2015).

tentions are ultimately benign. Since all challengers understate the scope of their intentions, great powers should ignore costless promises (Schweller, 1994); especially when their costly military choices imply they are greedy (Katagiri & Min, 2019).

I offer a framework to explain shifting beliefs about a challenger's motives in repeated interactions where challengers face incentives to understate the scope of their motives. I explain (1) how challengers exploit costless messages early on to coordinate beliefs and induce concessions; and (2) the conditions under which great powers eventually grow mistrustful about the challenger's long-term motives. I make predictions about when, and in what cases, great powers should shift from cooperation to competition based on qualitative differences in the demands that challengers make and not the scope of these demands or rates of militarization.

My key insight is that all states care about specific issues depending on what motivates their foreign policy (Moravcsik, 1998). States motivated by ethnic-nationalism, for example, covet different concessions than states motivated by security, prestige or revenge (Jackson & Morelli, 2011). For example, China would (probably) not accept territorial control over Uganda instead of Taiwan. China's historical context and motives 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 two points. First, state motives can vary along two dimensions: *scope* (some want more than others) and *preference order* (some value different objectives). Second, there are a finite set of possible motives a state can hold, and each drives that state to desire a few, specific objectives. For example, there is no motive that implies China wants to conquer Taiwan and Uganda but nothing else. If China fought for Uganda, the US would infer that China will fight for other issues when the opportunity presents itself. The limited set of motives creates discontinuities in the number of issues that the challenger could want, and this provides

nuanced opportunities for signaling across time.

When a challenger first starts making demands, types with expansive aims want to understate the scope of their motives to avoid competition for as long as they can. But, similar to the mechanism identified by Trager (2011), all types want to coordinate on preference order to receive valuable concessions first. In past theories, incentives to understate scope and coordinate preference order varied independently. In my theory, the challenger's long-term goals are tied to a specific motive that implies each type cares about specific concessions. When challengers justify their initial demands as in service of a specific motive, they reveal information about their scope and preference order simultaneously. Great powers could dismiss these early promises as cheap-talk. However, they prefer to exploit them as a benchmark to evaluate all future behavior against. From then on, they analyze if the challenger's behavior is consistent with her initial justification to determine if her long-run intentions are limited.

How the challenger justifies her demands at the beginning has important implications for patterns of cooperation and competition later on. Each motive implies that the challenger cares about only a few, specific territories. Once the challenger has captured all of these territories, she either makes another demand and reveals her initial justification was dishonest, or accepts the status quo forever. Types with expansive aims make another demand and great powers infer that their motives are vast. It is this revelation that triggers competition.

I develop a new framework to formalize uncertainty about intentions in world politics. I integrate studies that recognize diverse motives drive conflict in specific cases (Kaysen, 1990; Schultz, 2017; Shelef, 2016; Carter & Goemans, 2011) into broader theories of reassurance (Kydd, 2005; Powell, 1996; Yoder, 2019), 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, & 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 challengers demand, and explain how costly signals and costless diplomacy work together.

Section 1 develops an informal theory of state-motives. Section 2 integrates that theory of state-motives into a formal model of repeated concessions under the reassurance dilemma. In each section, I will evaluate my theory of state-motives and its strategic implications using archival case research from Anglo-Soviet relations (1940-1946).

1 A Theory of Intentions

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 this theory, principles refer to the underlying motives that states hope to satisfy through their foreign policies (Jackson & Morelli, 2011). Scholars often theorize about one of these principles at a time. States sometimes fight to unify their ethnic group (Goemans & Schultz, 2013), restore historical borders (Carter & 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 set of principles 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 an challenger wants based on what principle motivates it. This is different from the demands an challenger 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.

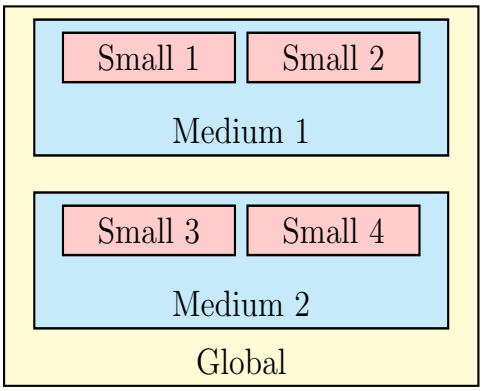
In practice, there is some ambiguity about whether a few specific issues are consistent with an challenger's principles or not. However, in most cases there is usually many more issues that clearly fit or do not fit. For example, it was clear that Africa, Asia, and the Americas did not fit into Hitler's nationalist aims but that parts of Austria, France and Czechoslovakia did.[6]

By distinguishing between the objectives that a challenger seeks and the principles that motivate its foreign policy, I assume a different set of possible challenger-types than what reassurance scholars typically study. To highlight these differences, I depict a stylized version

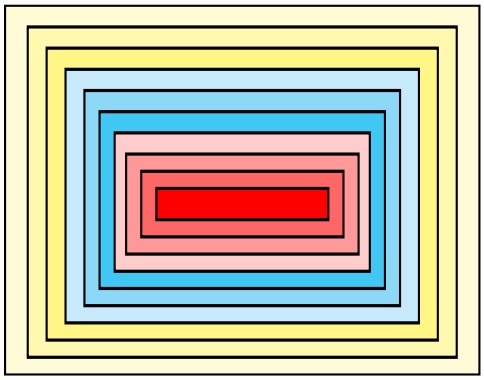
[6]Including some ambiguous issues does not change my predictions because great powers treat these ambiguous issues as if they are connected to each principle.

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.

of my type-space in Figure 1(a) and the conventional wisdom in Figure 1(b).[7] Think of the total area as a map that represents all the territories an challenger 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 challengers 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 scope.

Second, standard accounts assume that there is near-continuous variation in the number of issues that a challenger can care about. This drives the logic of salami-slicing in the reassurance literature: challengers repeatedly make a new demand and promise it will be their last (Jervis, 1978). This promise is plausible because there is always a type that values just one more concession. But it is not credible because types that want many more

[7] Press (2007) offers a more extreme alternative that I will also opperationalize later. But it is too difficult to draw in a picture.

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 challenger to want just one more territory. When an challenger makes one demand, it usually implies she wants, at minimum, several concessions.

The limited variation I describe explains why we do not observe challengers 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 exists that implied Monroe cared for 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.[8]

I draw two expectations that emphasize how my theory of principles departs from typical assumptions about state-motives:

Expectation 1 *Great powers are uncertain about which principle motivates their rival. They believe their rival's may be motivated by different principles. Each principle implies a discrete set of objectives.*

Expectation 2 *Great powers evaluate their rival's long-term demands by theorizing about different principles that may motivate them. They do not discuss the scope of their rival's demands independent of an underlying principle.*

[8]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.

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

From 1940 onward, the British War Cabinet exerted "a continual attempt to interpret Russia policy, assess the real intentions of Stalin and the small governing oligarchy of the USSR."[9] 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, and otherwise solicited theories from the military and Foreign Office.

Between 1940 and 1942, analysts reported only five different arguments about what Stalin's post-war demands could have been. Consistent with my theory of principles, each arguments was based on knowledge of Soviet history, and theories about the different principles that could motivate Stalin. Advisers stationed in Moscow emphasizes that Stalin was motivated by fear from foreign invasion. They argued that 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."[10]

Others still believed Stalin's interests would converge with Tsarist Russia's historical ambitions of power politics in Asia.[11] As a result, Stalin would seek territory in Iran, the Caucuses, the Baltic States, Central Asia and, possibly, India.[12]

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."[13]

Finally, a group believed that the spread of global communism motivated Russian foreign

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

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 wherever possible.

Consistent with my theory, each analyst argued that Stalin would want to capture multiple new territories once Hitler was defeated. But each thought that Stalin would seek different territories. There is some overlap across some of these claims and also variation between them. For example, if Stalin went after Iran he was either motivated by spreading communism or reverting to Tsarist ambitions and not motivated by commercial or security interests. In contrast, an ice-free port in Norway, only fit with commercial interests. Furthermore, there were large disjunctures in the scope of these demands. Only the spread of communism would drive Stalin toward global ambitions. But if Stalin wanted less than that, his demands were confined to a handful of specific demands.

2 A principle-driven theory of reassurance

I build a strategic model where a challenger's motives vary because she is motivated by different principles. These principles, imply that a challenger values specific territories. There are a finite number of principles that may motivate a challenger and each drives her to value specific foreign policy objectives. When I say that a great power is uncertain about an challenger's type, I mean that he does not know what principle motivate the challenger's foreign policy. But if he did know, he would know what concessions the challenger would demand both now and in the long-run.

First, I motivate a new game-theoretic framework to study heterogeneous motives in the shadow of repeated demands. Second, I set-up my model and opperationalize a type-space that matches my theory of principles. Third, I solve for an informative equilibrium. Fourth, I contrast my results with the existing theories of reassurance to develop unique predictions.

2.1 Informal description of strategic setting

I study an interaction between a great power, who wants to maintain the status quo, and a challenger, who has repeated opportunities to revise the status quo. As we shall see, whether the challenger profits from revising the status quo depends on her intentions. Recently, researchers have focused on different reasons that challengers get repeated opportunities to make demands including power transitions (Powell, 1999), technology change (Debs & Monteiro, 2014), a guns-butter trade-off (Leventolu & Slantchev, 2007), or shifting regional priorities (Treisman, 2004). They use the spatial bargaining framework to precisely link these reasons for repeated demands to both patterns of concessions and the causes of war. We now know that offers must reflect the distribution of power and interest; and that the fear of large concessions heighten incentives for competition between perfectly informed states.

But the bargaining framework is poorly suited to study uncertainty about an challenger's long-term motives because short-term bargaining incentives drive challengers to overstate, not understate their long-term intentions (cf Bils & Spaniel, 2016). When these models introduce incomplete information, war comes because great powers make low-ball offers that greedy (or highly resolved) challengers are unwilling to accept (Reed, 2003; Schultz & Goemans, 2019). These dynamics may match a few crisis episodes (e.g. the Alaska Boundary Dispute where the United States claimed it would fight at any cost) but not wide-spread, strategic competition between great powers that dominate the historical record (where the United States assured the British that it's demands were confined to the Western Hemisphere) (cf Kennedy, 1989). Older research used different assumptions to study repeated demands where challengers want to understate their motives (Powell, 1996; Jervis, 1978). 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.[14]

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

[14](Kydd, 2005) considers a two-period model. I want a framework that last several periods.

rate each period unless the challenger chooses to stop the game and accept the status quo. I define both cooperation and competition in terms of a the great power's willingness to make a concession. The great 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 challenger can make. In my model, competition reflects the great power's broad foreign policy designed to prevent or hinder the challenger from making additional demands (Waltz, 1979). This is different from a choice to fight a local war designed to impose a settlement over a specific territory in an immediate crisis (Fearon, 1995). Strategic competition can take the form of major war designed to overthrow a government (a-la the Second World War), but also includes the forward deployment of forces designed to prevent an challenger from meddling in the affairs of third-party states (a-la US policy towards the Soviet Union), or economic sanctions designed to prevent an challenger from having the resources to dedicate to foreign policy issues (a-la US policy towards North Korea).

Following the bargaining literature, I make sure that players' value for competition and the status quo move at the same rate. This assumption guarantees the three core result of the bargaining model obtain. First, each period the challenger 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 challenger's preferences order. Economists (Jackson et al., 2015) and political scientists (Joseph, 2017) show that uncertainty about a state's relative values across multiple issues produces dynamics that one-dimensional models cannot capture (Fearon, 1995). However, the amount of information that can be conveyed is limited (Penn, Patty, & Gailmard, 2011) when states' values for each issue is drawn independently.

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. In it, each challenger-type values a specific configurations of issues that are close together in space. This creates opportunities for great powers to draw inferences about all the issues that an challenger cares about based on specific demands she makes. Later, I contrast my results with counter-factual type-spaces that opperationalize the assumptions of the existing literature.

2.2 Formal Set-up

I model a challenger (R) and great power (D) that bargain over an nxn matrix Q_t^D.[15] The game includes opportunities for signals and concessions across time. I describe an arbitrary period with a subscript t and specific periods with numeric subscripts.

Like Figure 1(a), 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 takes on a dichotomous value: $\{0, 1\}$. When the element equals 1, D controls that element, when it equals 0, R controls that element. I 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. Each period D is forced to transfer one entry in Q_t^D to R.

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 same entries from matrix A and B. For example:

$$\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix} \times \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_2y_2 \\ x_3y_3 & x_4y_4 \end{bmatrix}$$

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

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

The core claim in my theory is that the type-space has important implications for peace, competition and information transmission. As a result, I assume a type space similar to Figure 1. 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.[16]

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.

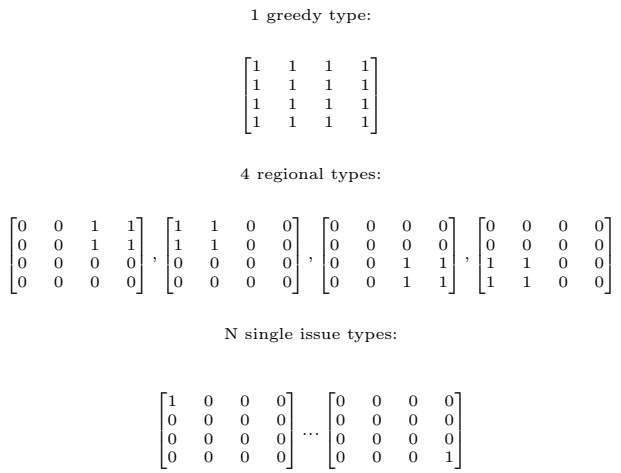
It is sometimes convenient to refer to sub-types and super-types. Any type ω_j is a sub-type (or nested type) of ω_i if all the elements that the former values so does the latter. I use set notation for sub-types and say: $\omega_j \subset \omega_i$ if $f(\omega_j \times \omega_i) = f(\omega_j) < f(\omega_i)$. 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 $\omega_i \supset_m \omega_j$).[17]

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 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

[16] I assume R values core and peripheral interests 1 and 0 respectively. In Appendix C.2 I assume R values peripheral interests ϵ and obtain the same predictions.

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

Figure 2: Construction of types in the 4x4 game



stop the game against the prior beliefs about types and the history of signals and offers.[18]

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.[19]

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.

   (a) If D transfers an element, D chooses one element to transfer and the game repeats.

[18] 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.

[19] 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.

(b) If D stops the game, the game ends and competition payoffs are realized.

Think of R stopping the game as if the challenger stops accepts the status quo. It may reflect a challenger 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: $r_R | U^R(s_D, V, y())$.

Finally, I make two assumptions to produce the re-assurance problem I want to study:

$$\mathcal{A}_1 : 1 < f(P) + c < 3N/4.$$

The right-most inequality implies D's cost of competition is less than 3/4 of D's total value for foreign policy issues. When this assumption is violated, D never fights no matter what R's motives are. The left-most inequality ensures that D always prefers to make one more concession rather than compete. If this assumption is violated, D selects competition in the first period.

I also assume:

$$\mathcal{A}_2 : f(P) > c.$$

This guarantees R keeps making demands and faces competition if there are foreign policy issues that they value, but do not yet control. When this assumption is violated, all challengers, even the greediest ones, accept the status quo when the great power credibly promises competition.

2.3 My theoretical mechanism

I claim there is a Pure Bayesian Equilibrium (PBE) where challengers 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 great powers will use to learn about the scope of the challenger's motives.

When I say that this message induces early concessions, I mean that absent communication, the great power would select competition in the first period. To establish this claim, I first solve for a babbling equilibrium. As with all babbling equilibrium, R's message at any point of the power transition does not influence D's beliefs.

Lemma 2.1 *In the* ***babbling equilibrium****, if R demands even one concession, D responds with competition straight away. If R only values a small number of issues (*$\max(f(P \times V)) < c$ *is satisfied), she never demands a concession. Otherwise, R demands a first concession and faces competition.*

See Appendix A.1. This equilibrium reflects Mearsheimer (2001)'s tragedy of great power politics. If D knew R's motives, and R's motives were limited, D would prefer to concede R what she values. The problem is that D does not know what R values. When D observes R demand concessions, D responds with competition because he does not know which or how many issues he must concede.

The conventional wisdom is that challengers cannot communicate information about their long-term intentions because they face insurmountable incentives to under-state the scope of their demands (Waltz, 1979). As a result, the babbling equilibrium should be unique. I now solve for an equilibrium where rising and great powers exploit costless diplomatic messages to coordinate their preferences leading to more efficient outcomes.

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 $\sigma(\hat{\omega})$. 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.[20]

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 $t = T_1^*$.[21] They all promise that they are the next-largest type. For a technical description of this message see Appendix A.3.

As with many cheap-talk models, there are multiple messages that can form a PBE. I chose this message for two reasons. First, it produces an informative equilibrium that sur-

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

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

vives the D1 refinement. Second, all of the other informative equilibrium that survive D1 also include a credible first period message that coordinates on preference order, dictates the sequence of offers that follow, and ensure that competition only comes if R does not stop the game in period $t = T_1^*$. But other messages produce equilibrium with many intermediate steps. I am careful to draw predictions based on similarities between these different informative messages. However, I describe the simplest message.[22]

Proposition 2.2 *Given assumptions $\mathcal{A}_1, \mathcal{A}_2$ there is always an **informative equilibrium** where R sends message $y(\sigma(\hat{V} \leq T^*))$. The threshold T_1^* emerges endogenously based on the cost of war and D's prior beliefs. In equilibrium, there are three pathways the game can take:*

1. *If R values less than T_1^* issues ($f(V) < T_1^*$) then in the first period R sends an honest message, and D updates his beliefs such that $pr(V = \hat{V}|\sigma_1(\hat{V})) = 1$. D offers R all of the issues that $\hat{V}$ values. Once there are no more issues that $\hat{V}$ values, R stops the game and accepts the status quo.*

2. *If R values exactly T_1^* issues ($f(V) = T_1^*$) then in the first period R sends an honest message, and D adjusts his beliefs about R's type. D is now more confident that R wants exactly what she said she wants, but D also believes it is possible that R understated her motives. D offers R the T_1^* issues that $\hat{V}$ values. Once there are no more issues that $\hat{V}$ values, R stops the game and accepts the status quo.*

3. *If R values more than T_1^* issues ($f(V) > T_1^*$) then in the first period R sends a dishonest message that implies R values T_1^* issues and is a sub-type of R's true type. D updates his beliefs such that D is most confident that R's message was honest, but D also believes it is possible that R understated her motives. D offers R the T_1^* issues that $\hat{V}$*

[22]There is another plausible message where all types send a first period message $f(\hat{V}) = T_1^*$ issues. That is, types with less than T_1^* valuable issues overstate their motives. This message expands the conditions under which my result applies because (1) it raises D's beliefs that R will stop making demands in period $t = T_1^*$; and (2) allows R to communicate more information about the scope of her demands with a first-period costless messages designed to reveal preference order. I do not focus on this message because it reveals much more information about the scope of R's demands than other informative equilibrium. I chose the simplest message that shares common features of other informative equilibrium.

values. Once there are no more issues that $\hat{V}$ values, R makes an additional demand. D responds with competition.

The mathematical description of the equilibrium and proof are in Appendix A.3 and a numeric example in Appendix A.4. Below I provide an intuition for two surprising facts about this equilibrium. First, R's first period message is so informative that D is willing to concede everything that R claims to value and test R's claim, rather than compete in the first period. Second, D prefers to switch to competition in period $t = T_1^*$ once R reveals that her first-period message was dishonest because R's second message is not informative even though R's first message was.

R's first period message is informative because all types want to coordinate over their preference order but only types that value more than T_1^* issues want to under-state the scope of their demands. To understand R's incentives to send different messages, suppose that there is a threshold T_1^* such that if R promises to want T_1^* or fewer concessions, D is willing to concede exactly T_1^* territories.

If T_1^* exists, then challengers that value T_1^* or less have no incentive to understate their scope. However, they will only receive offers they value if they honestly reveal issues they care about. They can achieve this by sending an honest message: $\hat{V} = V$. Out of all the types that send an honest message, there is a special group that value exactly what D is willing to offer: $f(V) = T_1^*$. These type faces the same incentives as the types with fewer interests. However, greedier types pool with this signal.

Types that value more than T_1^* issues face competition in the first period if they signal honestly. These types want to both: avoid competition for as long as possible; and induce valuable concessions. To avoid competition for as long as possible, they justify their first period demands with a message that implies they care about exactly T_1^* issues: $f(\hat{V}) = T_1^*$. To receive valuable concessions, they appeal to a type that is nested within their true type ($\hat{V} \subset V$). Since these types care about many issues, there is more than one different

principle they can appeal to that values exactly T_1^* issues and is nested within them.[23] The more expansive R's true aims are, the more lies it can tell to justify her initial demands and still gather T_1^* valuable concessions.

Even though greedier types will pool with limited aims types that value exactly T_1^* issues, R's message increases D's confidence that R wants what she says she wants relative to being a greedier type for two reasons. First, D rules out the possibility that R is any type other than $\hat{V}$ and all super-types of $\hat{V}$.[24]

Second, of these remaining types, each has a different probability of sending the message $\sigma_1(\hat{V})$. The type $V = \hat{V}$ always sends an honest messages. Since the type $V = \hat{V}$ 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 lies that they can tell.

[23] In practice, Hitler appealed to nationalism, but really cared about dominance through Eurasia. His nationalist goals intersected with his true aims. But he could have also appealed to security, revenge against Russia for the First World War, etc because all of these claims would have led him to take things he cared about.

[24] This includes types who care about more than T_1^* issues but whose interests do not intersect with $\hat{V}$.

We are now ready to consider D's incentives to make offers at different stages of the game. In the first period, D prefers to make concessions for T_1^* periods and then switch to competition only if he discovers that R has lied to him, rather than compete straight away if:

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

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

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. In equilibrium, R stops if her message was honest and keeps going if R understated her motives in the first period. In Appendix A.3, I shown that for any matrix of size N, D's posterior beliefs imply $\lambda_{T1} > .75$. Given assumption $\mathcal{A}_1$, I can always find a T_1^* that satisfies this inequality.

It follows, that D is willing to make exactly T_1^* concessions if R sends a first period message that implies she cares about T_1^* issues, and D can credibly threaten to switch to competition at period $t = T_1^*$, rather than keep making concessions at T_1^*, or switch to competition in an earlier period. Holding D's beliefs about R's motives constant, D's incentives to compete are largest in the first period. Thus, if D prefers cooperation in the first period, he will keep making concessions unless he changes his beliefs about R's long-term intentions.

Period $t = T_1^*$ is the first period in which D can learn if R's first period message was honest. In this period, types that sent an honest first-period message have captured all of the issues that they care about and therefore have no incentive to keep making demands. Thus, T_1^* is a critical juncture because it is the point where the challenger cannot demand another concession without revealing that her initial message was dishonest.

Just because D discovers that R's original message was dishonest does not mean he is

willing to switch to competition. After all, D still pays a cost of competition, and there is some chance that R is only the next largest type. At period T_1^* D prefers to keep cooperating and hope R is the next largest type, rather than switch to competition if:

$$(N - 3T_1^*)(1 - \lambda_{T*+1}) + (N - 3T_1^* - c - f(P))\lambda_{T*+1} < (N - T_1^* - c - f(P)) \tag{9}$$

$$\equiv \frac{3T_1^*}{f(P) + c} > \lambda_{T*+1} \tag{10}$$

λ_{T*+1} is D's beliefs in period T_1^* that R is the next largest type given the message that R sent in period $t = T_1^*$ and the fact that R did not stop the game at $t = T_1^*$. If the condition holds, the equilibrium I describe is not satisfied because D cannot credibly threaten competition at $t = T_1^*$.

Even though inequality 8 and 10 take the same form, inequality 10 is more difficult to satisfy for two reasons. First, the left hand side is 3 times larger than the left hand side of inequality 8. The reason is that R's type depends on her underlying principles, and the next largest type values $4T_1^*$ concessions. Since D has already conceded T_1^* territories, D understands that if R does not stop the game that D can expect to make at minimum another $3T*_1$ concessions.[25]

This discontinuity in R's value does not necessarily doom R and D to competition. If R can send a message that increases D's confidence that she is the next largest type, she could avoid competition. But R's message in period T_1^* has no effect on D's beliefs. This is surprising because R's first period message was so informative that it induced T_1^* concessions. Why can't R send a second credible message that leads to even more concessions? The reason is that R's first message exploited R's incentive to coordinate on preference order. But once R sent this message, D rules out all variation in preference order because these types sent

[25]Recall this was a historically grounded feature of the type-space. There was large discontinuities between what Stalin would have wanted if he was motivated by security, or communism. The same is true for Hitler and nationalism, or the United States and control of the Western Hemisphere.

different messages. At period T_1^*, all of the feasible types want to promise that they are the next largest type and so all pool their message. It follows that R cannot send a second credible message because there is no variation in preference order remaining for R to exploit.

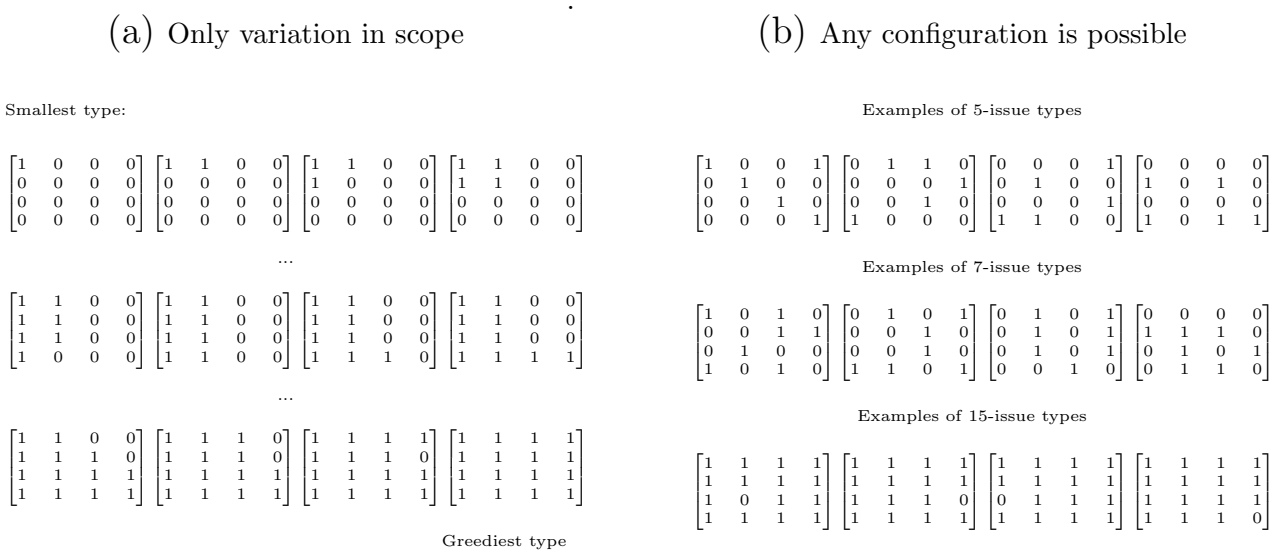
Appendix A.3 confirms that I can always find a T_1^* that implies D is willing to make first-period concessions if D can credibly promise to switch to competition at $t = T_1^*$ (inequality 8 is satisfied); and that D is willing to switch to competition in period $t = T_1^*$ (inequality 10 is not satisfied).

2.4 Unique predictions

I claim that equilibrium behavior described in proposition 2.2 is novel (i.e. differ from other re-assurance theories). I also claim these novelties arise because I assumed that the challenger was motivated by principles. In this section, I help validate that claim by adjusting the type-space to match the existing re-assurance literature holding all other features of the model constant. Figure 3 describes two counter-factual type-spaces that are commonly assumed in the existing literature. Figure 3(a) models the case where there is only variation in scope. Figure 3(b) models the case where states can value any configuration of issues. If I replace my type-space with either of these and hold other features constant, I cannot re-produce the equilibrium behavior described in proposition 2.2.

Contrasting my theory with these counter-factual type-spaces highlights that my theory produces new predictions at two critical junctures in repeated interactions. I call these critical junctures **the moment of focus** and the **moment of truth**. **The moment of focus** is the moment after the challenger first explains why her motives are limited. Before that moment, the great power is deeply uncertain about what the challenger wants. He believes that the challenger could be motivated by many different potential motives that each imply different long-term demands. After that moment, the great power focuses on whatever the challenger said she wants and rules out all other potential limited aims. The great power still believes that the challenger can hold more expansive aims. However, he is

Figure 3: Alternative type-spaces common in existing research



Each sub-Figure depicts a type-space that is consistent with assumptions made in the existing literature. In (a) there are N types in total and only variation in scope. In (b) there is one type for every possible combination of 0/1 values in the matrix. There are 65,674 possible types in the 4x4 matrix including 16 types that care about either 1 or 15 issues and 1 type that cares about 16 issues.

sufficiently optimistic that the her initial explanation was honest that he is willing to make concessions. Furthermore, I expect all future assessments to be framed around whether or not the challenger's initial message was honest.

These results do not obtain in equilibrium if R's motives only vary in scope (Figure 3(a)). In this case, R's first message is never informative and does not influence the strategic interaction because all types want to understate the scope of their demands and there is no variation in preference order to coordinate on.

These results also do not obtain in equilibrium if R can value any configuration of issues (Figure 3(b)). In this case, R can sometimes send an informative message in the first period. However, that message produces very different strategic behavior than the mechanism I describe. Each period, R sends a new message that implies she values just one more issue. Since there is complete variation in preference order, D is unable to draw any inferences about what issues R will demand in the future or how many more demands R will make based on R's past choices. As a result, R's message only reveals the next issue that R will

demand, and does not explain the set of issues that R will want over the course of the power transition.[26]

The following predictions highlight these differences at the moment of focus:

Expectation 3 *Great powers increase their confidence that they understand what a challenger's strategic intentions are after the challenger communicates it's underlying principles for the first time.*

Expectation 4 *Once the challenger has explained its principles great powers set out to validate if the challenger's declared principle is honest.*

The moment of truth follows the point where the challenger has taken all the concessions that are consistent with her original promise and must choose between remaining satisfied with the status quo, or exposing that her long-run intentions are greater than what she originally promised.

This period is critical because it present the first opportunity for the great power to learn if the challenger's initial justification was honest. If the great power discovers that the initial claim was dishonest, he becomes deeply mistrustful of the challenger. He may not know exactly what the challenger wants. However, he is sufficiently concerned that her aims are so expansive that he will change his strategy from cooperation to competition.

In contrast, in either of the counter-factual type-spaces I describe in Figure 3, D chooses competition in the first period or not at all. The reason is that all challengers face incentives to under-state their long-term motives and it is always possible for them to promise they want just one more demand. As a result, all greedier types pool on sending this sort of message. These incentives drive a salami-slicing logic which produces competition in the first period.

Expectation 5 *Great powers increase their confidence that a challenger's intentions are greedy after they witness behavior that is inconsistent with its early diplomatic promises.*

[26]This counter-factual emphasizes that R's first period message in my theory provides information about scope and preference order simultaneously because the type-space is constrained.

Expectation 6 *Great powers do not change their assessments about the challenger's intentions so long as the challenger's behavior (even violent behavior) is consistent with the challenger's first declared principle.*

2.5 Alternative explanations

Some argue that non-rational mechanisms cause great powers to update beliefs about the challenger's motives. Some predict that great power start out convinced that the challenger is aggressive but delay competition anyway (Schweller, 1994); others predict leaders succumb to biases that cause them to rely on personal experiences to form and change beliefs (Yarhi-Milo, 2014; Holmes, 2013). I do not have space to fully address these alternative, and I accept they may influence individual decision-makers to some extent. However, I chose predictions 3-6 to help distinguish my mechanism from these non-rational alternatives. My theory predicts great powers only change their beliefs at these two critical junctures. In the time in between, the great power's assessment of the challenger's long-term intentions should be consistent. It is good evidence that my mechanism works if there are lots of territorial demands, diplomatic meetings, or other key events between these two junctures and the great power's beliefs about the challenger's motives remain roughly constant.

Others argue that great powers start out very confident that challengers are greedy, or that challenger's can start out with limited aims and become greedy over time (Mearsheimer, 2001). If true, then challengers should not be able to convince great powers that their motives are limited, and great powers should choose competition at the first violent event.

Others argue that power parity (Organski & Kugler, 1980; Reed, 2003), or sudden shifts in the distribution of power (Powell, 1999) drives competition and not shifting beliefs about preferences. If this true then great powers should not think about the long-term motives of emerging threats, and changes in beliefs about motives should not be correlated with competition choices.

To be clear, I think that theories of competition based on power or shifting preferences

provide important scope conditions for my theory. In the Appendix I search for conditions when my mechanisms best explains patterns of competition and when these alternatives are better explanations. In C.1, I introduce shifting power and give both players an outside option of war in the form of a costly lottery (a-la Powell, 1999). In C.2 I assume that R values peripheral interests $L > 0$, and not 0. In C.3 I assume that R can start out with limited aims but become greedy with probability. In C.4 I assume D starts out with stronger priors that R is the greediest possible type.

These extensions show that when power shifts rapidly, or the great power holds prior beliefs that the challenger is likely to be (or become) greedy, or the challenger cares about peripheral interests almost as much as her core interests, the great power is unwilling to test the challenger's motives as my theory expects. In these extreme cases, the great power enacts competition following the challenger's first demand; and diplomacy has no reassuring effect.[27] However, in moderate parameter ranges, my motives-based logic dominates the strategic setting in every extension.[28] 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.

Although I cannot test every possible extension to my theory in one paper, I am optimistic that my predictions are robust to complex strategic settings and not an artifact of my bargaining assumptions. When I use the type-spaces that the existing literature assumes, my model's predictions conform to the existing literature's predictions. If my modeling assumptions biased my result, they should also bias result with these counter-factual type-spaces. They do not.

In Appendix B, I consider more complex type-spaces that are consistent with my theory. In each case, I describe equilibrium conditions similar to those described in proposition 2.2. The results illustrate that as the type-space grows in complexity in ways that may damage

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

[28] 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.

informative messages, R sends more complex messages that tie together scope and preference order in different ways to recover the basic result in proposition 2.2. The reason is that types who value more issues (increased scope) have more flexibility in the messages they send, and so can adjust how they provide information about their preference order to manipulate D's beliefs about scope.

2.6 British assessments of Soviet intentions (1940-1946)

I report archival evidence from Anglo-Soviet (1940-1946) relations to demonstrate that my mechanisms explains how the British form and update their beliefs about Soviet intentions. I focus my analysis on two different periods that correspond with the moment of focus (1940-1941) and the moment of truth (1945-1946).

2.6.1 The moment of focus

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 above, every analyst 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."[29] 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 the present stage of the war."[30] 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

[29]FO371/29472, N5679/3014/38
[30]Prem 3/395/6, Nov. 10, 1941

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."[31]

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."[32] 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 to permanently take 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."[33]

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."[34] Upon reading Eden's report, the War Cabinet tasked sub-committees to establish if Stalin truly was motivated by security.[35]

This evidence supports my third and fourth prediction about why private diplomacy is so

[31]Prem 3/395/6, Nov. 10, 1941
[32]Prem 3/394/3 Dec. 18, 1941
[33]WP(42)8; N109/5/38(1842).
[34]FO 371/32/876 Feb. 12, 1942
[35]See WP(42)69 ; WM(42)18 ; N798/5/38 ; N1024/5/38; N1279/5/38; N1526/5/38; T395/2/402 (Churchill Papers); N1395/5/38); N1653/5/38; WM(42)37, C.A. for extensive discussion about Stalin's intentions relating to security.

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.

2.6.2 The moment of clarity

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 challenger's domestic political discourse, cause great 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"[36] report was approved by the Foreign Office and the Chief of Staff Committee (COC) and sent to the War Cabinet and prime minister.[37]

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 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

[36] Goodman (2014)

[37] 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, & Goodman (2014).

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.[38] 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."[39]

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..."[40] 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"[41] 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."[42] These calls for calm persuaded the balance of policy-makers including the Prime Minister and Foreign Minister.[43]

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

[38]He promised not to do this at Yalta.
[39]FO/371/50912 Lockhart's Minute, Apr. 11 1945
[40]Clark Kerr's opinion: FO 371/47076 Apr. 16, 1945
[41]Clark Kerr's opinion: FO 371/47076 Apr. 16, 1945
[42]Clark Kerr's assessment: FO371/47941 Mar. 27, 1945
[43]who concurred on this Memorandum.

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."[44] 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."[45]

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 Conclusion

Patterns of competition and cooperation have puzzled scholars of reassurance for a century. I develop an explanation based on complex variation in the challenger's motives rather than how states moderate the scope of their demands, rates of shifting power, power parity, and non-rational mechanisms. My key insight is that states are motivated by different principles and these principles imply that specific objectives are valuable. Challengers promise that once they achieve all the concessions related to that principle, they will peacefully integrate in the existing world order. Great powers use these reassurances to evaluate the challenger's behavior from then on. They observe if military spending, demands and domestic policies are

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

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

consistent with what was said. They trust challengers when behavior is consistent, but become alarmed when behavior is not. My assumptions about state-preferences matched how states explain their preferences across power transition cases since 1850, and my theoretical predictions match Anglo-Soviet relations.

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 competition and we ought to prepare accordingly.

My theory of preferences can be applied to other strategic interactions between great powers. Rather than assuming that a state's preference over many different issues can only vary in scope, or is drawn independently, researchers can get theoretical leverage by assuming that a state's value for different issues is tethered together by what motivates its foreign policy. As a result, we can draw inferences from which issues states do or do not fight for about which issues they will fight for in the future. Other questions that could profit from this framework include how the United States selects which terrorist organizations or nuclear aspirants to resist and which to ignore, why states join some institutions (including PTAs and alliances) but not others, when states decide to shirk on their commitments.

References

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

Axelrod, R., 1980. More Effective Choice in the Prisoner's Dilemma, *Journal of Conflict Resolution*, **24**(3), 379–403.

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

Bils, P. & Spaniel, W., 2016. Slow to Learn, *Journal of Conflict Resolution*.

Brooks, S., 1999. 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. & Goemans, H. E., 2011. The Making of the Territorial Order: New Borders and the Emergence of Interstate Conflict, *International Organization*, **65**(02), 275–309.

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

Copeland, D. C., 2015. *Economic interdependence and war*.

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

Fearon, J. D., 1995. Rationalist Explanations for War, *International Organization*, **49**(03), 379.

Gilpin, R., 1983. *War and Change in World Politics*, vol. 1983, Cambridge University Press, Cambridge.

Glaser, C. L., 2010. *Rational Theory of International Politics*, Princeton University Press, Princeton.

Goemans, H. E. & Schultz, K. A., 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.

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

Hopf, T., 1994. *Peripheral visions : deterrence theory and American foreign policy in the Third World, 1965-1990*, University of Michigan Press.

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

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

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

Joseph, M., 2017. A Little Bit of Cheap-Talk is a Dangerous Thing.

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

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

Kennedy, P. M., 1989. *The rise and fall of the great powers : economic change and military conflict from 1500 to 2000*, Vintage Books.

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

Kertzer, J. D., 2016. *Resolve in international politics*, Princeton University Press.

Kurizaki, S., 2007. 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 University Press, Princeton, N.J.; Woodstock.

Leventolu, B. & Slantchev, B. L., 2007. The armed peace: A punctuated equilibrium theory of war, *American Journal of Political Science*, **51**(4), 755–771.

Mearsheimer, J. J., 2001. *The Tragedy of Great Power Politics*, Norton, New York.

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

Organski, A. F. K. & Kugler, J., 1980. *The War Ledger*, University of Chicago Press, Chicago.

Penn, E. M., Patty, J. W., & Gailmard, S., 2011. Manipulation and Single-Peakedness: A General Result, *American Journal of Political Science*, **55**(2), 436–449.

Powell, R., 1996. 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 University Press, Princeton, N.J.

Press, D. G., 2007. *Calculating Credibility: How Leaders Assess Military Threats*, Cornell University Press.

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

Sartori, A. E., 2005. *Deterrence by diplomacy*, Princeton University Press.

Schultz, K. A., 2017. Mapping Interstate Territorial Conflict, *Journal of Conflict Resolution*, **61**(7), 1565–1590.

Schultz, K. A. & Goemans, H. E., 2019. Aims, claims, and the bargaining model of war, *International Theory*, pp. 1–31.

Schweller, R. L., 1994. Bandwagoning for Profit: Bringing the Revisionist State Back In, *International Security*, **19**(1), 72–107.

Shelef, N. G., 2016. Unequal Ground: Homelands and Conflict, *International Organization*, **70**(1), 33–63.

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

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

Treisman, D., 2004. Rational Appeasement, *International Organization*, **58**(02), 345–373.

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

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

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

Yoder, B. K., 2019. Retrenchment as a Screening Mechanism: Power Shifts, Strategic Withdrawal, and Credible Signals, *American Journal of Political Science*, **63**(1), 130–145.

A Model's Solution

In this section I solve for the solutions to the basic model reported in the paper.

A.1 Lemma 2.1: Babbling Equilibrium

Since R mixes over all messages in every period there are no off-path messages to consider and D cannot be informed by any message. Thus, I focus only on stopping rules and D's offering rule.

Assume that R only stops the game once D has offered R all valuable issues. I'll show that D either stops the game in the first period, or after all issues are conceded.

Holding R's stopping rule fixed, D's best offering strategy is the one that most quickly rules out types. This strategy starts by randomly choosing a single issue that satisfies a single issue type ω_x, then offering the three other issues that ω_{x+1} values (ω_x's next largest super-type). Once D offers all issues that ω_{x+1} values, D moves to another 4-issue type $\omega_{y+1} \subset \omega_{x+2}$. That is another four-issue type nested within ω_{x+1}'s next-largest super-type and offers those four concessions in random order. D then completes this process for the four 4-issue types nested within ω_{x+2}. D then moves out to ω_{x+3}.

Define a sequence $K : \{k_t\}$ as the number of types that are ruled out each period under D's best offering strategy. Here subscript t is the period, and k is the number of types ruled out in that period. The sequence is: 1,1,1,2,1,1,1,2,1,1,1,2,1,1,1,3,1,1,1,2, 1,1,1,2,1,1,1,2,1,1,1,3,1,1,1,2,1,1,1,2,1,1,1,2,1,1,1,3,1,1,1,2,1,1,1,2,1,1,1,2,1,1,1,4, 1,1,1,2,1,1,1,2,1,1,1,2,1,1,1, 2,1,1,1,2,1,1,1,3,1, 1,1,2,1,1,1,2, 1,1,1,2,1,1,1,3, 1,1,1,2,1,1,1,2,1,1,1,2,1,1,1,4...

Notice that $\sum_{t=1}^{N} k_t$ is the total number of types w.[46] Also notice that no matter the size of the matrix, the first two elements in the sequence must be $1, 1$.

I'll now show that if D prefers to delay competition in the first period and instead make an offer, D never prefers competition. In the first period, D selects competition if $\frac{\sum_{t=1}^{N} k_t(N-t)}{w} < N - f(P) - c - 1$. But in the second period D selects competition if $\frac{\sum_{t=2}^{N} k_t(N-t)}{w-1} < N -$

[46] Strictly w is the space. Here I abuse notation and use it to mean the total number of types.

$f(P)-c-2$. Making the RHS of both inequalities $N-f(P)-c-1$ and separating k_1 out, we see that D's first period incentives are larger if: $\frac{\sum_{t=2}^{N} k_t(N-t)}{w} + \frac{(N-1)}{w} < \frac{\sum_{t=2}^{N} k_t(N-t)}{w-1} + 1$. Clearly, $\frac{\sum_{t=2}^{N} k_t(N-t)}{w} < \frac{\sum_{t=2}^{N} k_t(N-t)}{w-1}$ and $\frac{(N-1)}{w} < 1$. It can be shown by induction that D's incentives to wait only increase each period. It follows that D will either stop the game in the first period or not at all if R will not stop the game unless R has all valuable issues.

Turning to R's incentives to set a stopping rule. Consider the case that D will not stop the game. Clearly, R's best reply is to wait until she receives all valuable concessions. Consider the case where D will stop the game in the first period if R does not. R prefers not to stop the game and accept competition when: $\max(f(P \times V)) - c > 0$. Otherwise, R stops the game in the first period as written in the equilibrium.

The exact cost and risk that determines whether D is willing to make no offers or all offers depends on the size of the matrix. D prefers competition in the first period if $N - \frac{\sum_{t=1}^{N} k_t(N-t)}{w} < f(P) + c$. For a matrix size $N = 4$, prefers competition in the first period if $6/5 > 4 - f(P) - c \equiv f(P) + c > 3\frac{4}{5}$. Taking $N \to \infty$ we get $f(P) + c > 3(N-1)/4$ (from above) as stated in the Lemma. This implies that (approximately) D will stop the game in the first period if D's cost of war is less than D's value for 3/4 of all foreign policy issues.

A.2 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.[47] 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 $\sigma_1(\hat{V}_1)$

[47] 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 & 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.

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 $\sigma_1(\hat{V}_1)$.[48]

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.3 Proposition 2.2: 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.4 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. it also describes specific bounds on the equilibrium space and does not appeal to assumptions $\mathcal{A}_1, \mathcal{A}_2$.

[48] 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.

Proposition A.1 *Suppose*

$$\frac{T_{\hat{V}_i}}{f(P)+c} \geq \lambda_{T1} \tag{11}$$

and

$$3T_{\hat{V}_1} > c \tag{12}$$

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* $t = T_{\hat{V}_i}$

$$\frac{3T_{\hat{V}_i}}{f(P)+c} \leq \lambda_{T*+1}, \tag{13}$$

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 11 and 13 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 12).

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 11 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:

$$\lambda_1(f(Q_1^D) - T_1^*) + (1 - \lambda_1)(f(Q_1^D) - T_1^* - f(P) - c) > f(Q_1^D) - f(P) - c \equiv \frac{T_1^*}{f(P) + c} < \lambda_{T*1}, \tag{14}$$

Inequality 14 assumes that D prefers to compete in period $t = T_1^*$ if R does not stop the game. As a result, inequality 14 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:

$$\lambda_{T^*+1}(Q_0 - 4T^*) + (1 - \lambda_{T^*+1})(Q_0 - 4T^* - f(P) - c) \geq Q_0 - T - f(P) - c \equiv \frac{3T^*}{f(P) + c} \geq \lambda_{T^*+1} \tag{15}$$

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 ($\hat{V} \nsubseteq V$), 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 14. 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.[49] 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:

$$\beta_1(V = \hat{V} | \sigma_1(\hat{V})) = \frac{P(\hat{V})}{P(V = \hat{V} | \Omega, y) + P(V \neq \hat{V} | \Omega, y)} \tag{16}$$

[49] 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.

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

$$\beta_1(V = \omega_{+m}|\sigma_1(\hat{V})) = \frac{P(\hat{V})}{4^m * (P(V = \hat{V}|\Omega, y) + P(V \neq \hat{V}|\Omega, y))} \tag{17}$$

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

$$\lambda_{T*}(|f(\hat{V}) = T_1^*) = \frac{4^{\log_4 N - \log_4 T_1^*}}{\sum_{m=0}^{\log_4 N - \log_4 T_1^*} 4^m} \tag{18}$$

Given that all types $f(\hat{V}) > T_1^*$ mix in this way,[50] 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 14, if $\frac{T_1}{f(P)+c} < .75$ can be solved for some $T_1 \in f(w\,\{\omega_i\})$, then inequality 14 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 14 can be satisfied for any matrix size nxn.[51]

Inequality 14 assumes that D will not stop the game for periods $t < T_1^*$.[52] 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 15 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.[53] Types $\hat{V}_1 \subset V$ do not stop the game and

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

[51] With the smallest matrix (4x4), I can satisfy inequality 14 with competition parameters: $f(P) = 1$ and $c = 1/4$. It is easier to satisfy as matrices grow.

[52] 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 $t = T_1^*$.

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

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:

$$\lambda_{T*+1}(|f(\hat{V}) = T_1^*, r_R > T_1^*) = \frac{4^{\log_4 N - \log_4 T_1^* - 1}}{\sum_{m=0}^{\log_4 N - \log_4 T_1^* - 1} 4^m} \tag{19}$$

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).[54]

Plugging in $\lambda_{T*+1} = 1$ into inequality 15, the outer bound of the inequality is $\frac{3T_1}{f(P)+c} > 1$.[55] 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 14 and 15 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. In the manuscript, this is always true given $\mathcal{A}_1$.

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 14 and 15.

I'll now show that neither player can profit from deviating from the equilibrium described in proposition 2.2 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 14 and 15 D cannot profit from setting an alternative stopping rule. So I focus on D's incentives to alter his offering strategy.

[54] 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.

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

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 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 12 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.[56] 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

[56] 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.

some period $t \leq T_1^*$.[57] 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 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 14, 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 14 and 15.[58]

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.4 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 1.

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

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

Table 1: 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 $c = 3$, $f(P) = 10$.

To be clear, in the baseline model I assume that each of these types is equally likely to be drawn. In Appendix C.4 I show that even if I re-adjust the probabilities so that it is much more likely R is the greediest possible type, I can still produce an informative equilibrium.

Throughout the example, I refer to four possible realizations of R's type, which I depict in Figure 4. 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

Figure 4: Different types of R in the numeric example

V_s:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

V_m:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

V_r:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\overline{\overline{V_m}}$:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

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

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:

$$(64 - T_1^*)(1 - \lambda_{T1}) + (64 - T_1^* - 5 - 8)(\lambda_{T1}) > (64 - 5 - 8)\lambda_{T1} > \frac{T_1^*}{13} \tag{20}$$

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}).[59] 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:

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

$$(64 - 3T_1^*)(1 - \lambda_{T*+1}) + (64 - 3T_1^* - 5 - 8)\lambda_{T*+1} < (64 - T_1^* - 5 - 8) \tag{21}$$

$$\frac{3T_1^*}{13} > \lambda_{T*+1} \tag{22}$$

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 20 and 22 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.

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.[60]

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

$$pr(V = V_m|\sigma_1(V_m)) = \frac{pr(V = V_m) * pr(\sigma_1(V_m)|V = V_m)}{pr(V = V_m) * pr(\sigma_1(V_m)|V = V_m) + pr(V = V_r) * pr(\sigma_1(V_m)|V = V_r) + pr(V = V_G) * pr(\sigma_1(V_m)|V = V_G)} \quad (23)$$

$$\frac{1/85 * 1}{1/85 * 1/16 + 1/85 * 1/4 + 1/85 * 1} = 16/21 \quad (24)$$

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

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

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 20, 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 20, I assumed that D would stop the game if he discovered at T_1^* that R's initial message was dishonest. This implied inequality 22 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.[61] 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 22 is satisfied and D prefers competition to making additional offers at T_1^*.

[61] 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 Realistic adjustments to the type-space

I developed a stylistic type-space to operationalize my theory of motives as principles. That type-space included a specific amount of overlapping between types. In the real world, there is likely to be more complex overlapping between types. In this section, I describe the results if I adjust my type-space in three different ways that would still be consistent with my theory. I'll then describe R's messaging strategy in each case that produces the same equilibrium predictions.

Unfortunately, the model is too complex for me to derive analytical values that identify exactly how much overlapping the equilibrium in proposition 2.2 can withstand before it falls apart. What I can show through examples is that it does survive some overlapping, and (in the main paper) that it fails if there is only variation in scope, or no relationship between a state's principles and the objectives that she seeks. Thus, I can say that my result requires some overlapping and nesting of types, and also discontinuities in the scope of the demands of different types. However, it does not only follow given the specific amount of overlapping that I study in the manuscript.

These extensions all produce an interesting additional feature that is obscured in the baseline model. They each show that as the type space gets more complex, that the rising power must send more complicated messages (in terms of how greedier types mix) to produce the same strategic behavior. Interestingly, each of these messages only works because the rising power is able to send information about scope and preference order simultaneously. These features are partially obscured in the baseline model because the rising power is equally likely to be every type, and there is restricted overlap between types. As I relax these assumptions the messages become even more informative.

B.1 Overlapping Types of the same size

My type-space assumed that types who valued the same number of issues always cared about different issues. For example, for any two types that cared about 4 issues only, they both cared about different issues. In practice, there is likely to be more overlap between types that care about the same number of issues. For example, according to the British analysis, Stalin would have wanted about the same number of concessions if he was motivated by security or access to ports for commercial ends. In both cases, he would have sought control over the Baltic states.

My informative equilibrium survives in a world where types overlap. To demonstrate this, I extend the 8x8 numerical example above with competition parameters $P = 10, c = 3$ to include to include one additional type:

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This type values 16 squares and its interests overlap with two other types that care about 16 issues. In the numeric example used in the text, this new type overlaps with the true type V.

With this additional type, the general behavior described in proposition 2.2 still constitutes an equilibrium. There is the same critical threshold ($T_1^* = 4$), and D and R still play the same stopping rules and offering rules. There is a difference in how the greediest type mixes over the different messages that she could send. In particular, the greediest type does not mix evenly over all feasible four-issue types. Rather, she mixes unevenly to account for the overlap. Given R's different mixing, D has slightly different posterior beliefs.[62]

The following matrix helps explain how the greediest type now mixes. Each entry in the matrix represents a different 4-issue type. (and thus 4 elements in the 8X8 matrix). Each of

[62]The differences are small enough so they do not alter strategic behavior.

Period	Message observes	Pr. R wants 4	16	64
1	A	40/53	10/53	3/53
$T_1^* = 4$	A		10/13	3/13
1	B	40/63	20/63	3/53
$T_1^* = 4$	B		10/13	3/13
1	C	20/26	5/26	1/26
$T_1^* = 4$	C		5/6	1/6
Prior		16/86	5/86	1/86

these is a feasible message that the greediest type could send. I have categorized these types A, B, C based on how the greediest type mixes, and D's posterior beliefs depending on the message D observes.

$$\begin{bmatrix} A & B & B & A \\ A & B & B & A \\ C & C & C & C \\ C & C & C & C \end{bmatrix}$$

That greediest type, mixes over the possible four-issue messages in the following ratios:

$$\begin{bmatrix} 3/40 & 3/40 & 3/40 & 3/40 \\ 3/40 & 3/40 & 3/40 & 3/40 \\ 1/20 & 1/20 & 1/20 & 1/20 \\ 1/20 & 1/20 & 1/20 & 1/20 \end{bmatrix}$$

In period $T = T_1^*$, if the greediest type's first period message was a type A or C it claims to be the next (and unique) largest super-type. If the greediest type's first period message was a type B, then it has two different messages it could send and still promise to be the next largest type. It mixes evenly over both of them.

While D plays the same strategy, her beliefs are slightly different at different stages of the game depending on the message she observes.

These differences emphasize exactly how R overcomes more complexity in the type-space. Types that have more flexibility in the message that they send (here the greediest type) choose to mix in different ratios in the first period. This is possible because variation in scope also implies that types with the greatest incentives to under-state scope also have the most flexibility in the lies that they can tell. In this way, greedier types can exploit that flexibility to send messages that induce the same strategic behavior.

B.2 Natural resource deposits

IN the real world, some territories are so valuable that all rising powers, no matter their true principles will want them. We might think, for example, that no state in the Middle East would turn down absolute control of Saudi Arabia's oil fields no matter what principle motivated their foreign policy.

To model this sort of variation, consider the 8x8 numeric example above. But now assume that every type values the element $\{8, 8\} = 1$ in addition to what they already valued. As a result, all types that did not already value this element now value it. This implies that there is only 1 type that values 1 issue (the type that originally valued $\{8, 8\}$) and 15 types that care about 2 issues. Also, there are 3 types that care about 17 issues and still one that cares about 16 issues.

As a result of this difference, I searched for a slightly different set of behaviors that have the same substantive interpretation and produce the same predictions. In general, there is still a threshold that exists for the 4-issue types. But how the game plays out depends on exactly which issues R cares about. Define three different kinds of 4-issue types: A,B,C as follows:

$$\begin{bmatrix} A & A & A & A \\ A & A & A & A \\ A & A & B & B \\ A & A & B & C \end{bmatrix}$$

D's threshold is now also conditional on R's first period message. If R sends a message that implies she is a four issue type A or B, D sets the threshold at $T_1^* = 5$. The reason is that all of these types care about 5 issues not 4 (they care about the issues they identify, plus the entry $\{8, 8\}$). If R sends a message that implies she is a four-issue type C, D sets a threshold $T_1^* = 4$ because this type only cares about four issues.

Clearly, this implies that types who want to understate their motives but maximize the number of concessions that they will receive avoid sending a message that they are type C. But this does not effect all types $f(V) > T_1^*$. It only effects (1) the greediest type, and (2) the type that cares about the 16 issues in the bottom right corner of the matrix.

I now describe how these two types choose their first period message to induce an equilibrium behavior as in proposition 2.2.[63] The greediest type selects a message at random that she is one of the following 5-issue types with the following probabilities:

$$\begin{bmatrix} 1/20 & 1/20 & 1/20 & 1/20 \\ 1/20 & 1/20 & 1/20 & 1/20 \\ 1/20 & 1/20 & 2/15 & 2/15 \\ 1/20 & 1/20 & 2/15 & 0 \end{bmatrix}$$

The type that cares about the 16 issues in the bottom right quarter of the matrix sends a 4-issue message in the first period with the following probabilities:

[63] With the difference that D sets $T_1^* = 4, 4 + 1|\hat{V}$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 0 \end{bmatrix}$$

Otherwise, all types send the same message I specified in the main model.

As expected, these two types now avoid sending a message that they care about the very bottom right four-issue type to guarantee they receive 5 (not 4) valuable concessions.

To make the equilibrium hold together, the greediest type must send a message she is a 4 issue-type B more often than a type A. This adjustment is necessary to make sure that D is willing to switch to competition in the 5th period. If the greediest type mixed evenly over 15 types (4-issue types A and B), then D could not credibly promise to switch to competition if R did not stop the game in the fifth period. The reason is that R would be very likely to value just 11 more concessions and not the entire matrix. This would force the equilibrium to unravel.

In this case, D would never be able to promise to switch to competition in period $T_1^* = 5$ and the equilibrium unravels. This again highlights how D sends messages that tie together information about scope and order simultaneously to make the equilibrium fit together.

C Robustness Checks: Power and shifting interests

In this section, I describe the results of several model extensions. Each extension I discuss introduces an additional element common in studies of power, bargaining and war into my framework. I omit these additional complications from the model in the manuscript because I wanted to focus on the core mechanism that produced my result. Although these additional complications are important in studies that locate the source of war during power transitions in power variables, and bargaining incentives, they are less important in my framework. The purpose of including them here is to show the conditions under which we might expect

power dynamics to dominate as a mechanism for competition and when we might expect motives-based dynamics to dominate as a mechanism for competition.

C.1 Including shifting power and war

The baseline model focused on how the threat of containment could produce communication and cooperation during power transitions. Yet many power transitions end in major war and not containment. It is not obvious that my mechanism can be supported in the shadow of war, when R's military capabilities increase across rounds and this shift in power alters R's expected value from war as the power transition continues. I'll now describe an extension that explicitly models relative power between states, and gives both actors an outside option of war. I'll assumes each element in the matrix is divisible, and that the settlements states reach and their value for war is a function of relative power between them.

I'll describe an equilibrium similar to the equilibrium in proposition 2.2 when I explicitly model shifting power and war.

Adjustments to the base-line I assume that the elements of Q are infinitely divisible. I allow any element $q^D\{m,n\} \in Q_t^D$ with coordinates m, n can take on values $[0,1]$. However, $q^D\{m,n\} + q^R\{m,n\} = 1$. As a result, it is still true that $Q_t^D + Q_t^R = \mathbf{1}$ (an nxn matrix of ones).

In existing theories of power transition, strategic behavior is driven by the relative military power between states (usually written as p_t). Relative power between R and D usually dictates the consequences of competition as well as the terms of feasible bargains that both states are willing to accept. Existing research focuses on how power shifts across time (usually power shifts are marked by Δ such that $p_t = p_{t-1} + \Delta$).

I now explicitly model shifts in power across rounds. Define relative military power between R and D at time t as p_t. I assume that the game starts where $p_0 = 0$, then R has opportunities to invest at a cost each period such that if R invests in her military in period

t, $p_t = \max[p_{t-1} + \Delta, 1]$.

To bring my theory in line with standard models, I assume that shifting power influences the strategic setting in three ways. First, I assume that R's military investment is costly. As we shall see, R pays a cost c_m (m for militarization) each time that she invests in the military.

Second, I replace D's option for containment with an option for major war. I now model competition as a major war between R and D. If either player selects war in period t, the bargaining stops, and both players enter a lottery that R wins with probability p_t and D wins with probability $1 - p_t$. Both players pay a cost c_w for fighting a war, the winner of war gets to control all the issues and territories under dispute and payoffs are realized. As a result, if either player selects war at time t the game stops and players realize expected utilities:

$$EU_t^R(war|\omega_i, t) = p_t f(V) - c_w - tc_m \tag{25}$$

$$EU_t^D(war|t) = (1 - p_t)N - c_w \tag{26}$$

where c_m is R's cost of militarization[64] and N is the total number of issues in the matrix. As in the baseline model, D values Q uniformly and so takes a value of 1 for controlling it all (but only wins with probability $1 - p_t$). In the baseline game, the consequences of containment were constant. Here the consequences of selecting war are influenced by the balance of power and therefore shift across periods. In each period, both players have the opportunity to choose war. As a result, R's expected gain from war increases over time as R becomes more powerful.

Third, I directly tie D's offer to relative power rather than assume D must concede one issue at a time to R. I now assume that at any point in time, the issues that a player controls

[64] I'll explain this more below

depends on a bargaining process where D offers R an amount of territory conditional on the distribution of power. Define X_t as an offer at time t. Where X_t is an $n \times n$ matrix where each element $X_t\{m,n\} \in [0,1]$. If R accepts X_t then the status quo $Q_t^R = X_t$. For simplicity, I assume that offers in any period must increase (or hold constant) R's share of each issue in Q over time. That is for any element $Q_{t-1}^R\{m,n\} \leq X_t\{m,n\}$.[65]

I assume that in any period D's offer must be equal to the distribution of power and so $p_t = f(X_t)$. To be clear, I fix the total amount of territory that D must concede in each round as a function of the distribution of power. However, D still gets to decide which territory he concedes. A fixed offer at size $f(X_t) = p_t$ has several nice features. First, it is always in the bargaining range for any cost of war. Second, there is an alternating offers bargaining protocol that can produce this outcome. Third, it follows a growing literature that finds that states have a preference for the fair division of surplus in international relations. Fourth, it builds on existing work that fixes the offer at p_t.[66]

If R stops the game and accepts the status quo, players realize pay-offs:

$$U_t^R(status\ quo|x_t,\omega_i) = f(Q_t^R \times V) - tc_m \tag{27}$$

$$U_t^D(status\ quo|x_t) = f(Q_t^D) \tag{28}$$

The new sequence of moves is as follows. 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 rounds where

[65]This assumption allows me to ignore bizarre offering strategies that appear off the path.

[66]The result that follows is robust to other protocols that fix the offer as a function of p_t. For example, an offer fixed at $p_t - c_w$ will produce nearly identical results. However, the result is not robust to a take-it-or-leave-it protocol. When D can make a take-it-or-leave-it offer, D tries hard to offer R her minimum demand. In the shadow of war, D tries hard to offer R enough to leave R indifferent with war. But when R has limited aims, R's minimum demand is lower because she values fewer territories. As a result, D low-balls types of R that hold limited aims by giving them a very small portion of the pie. This produces incentives for R to over-state her preferences to maximize the size of the offer that she receives. In effect, this would model a setting where Hitler openly claimed that he wanted to take over the world, and the British thought that Hitler may secretly hold more limited intentions than what he claimed. This obviously fits no power transition through history. Fixing the offer as a function of p allows me to study the case where rising powers under-state their aims.

in each round:

1. R sends a message about it's type, and then either decides to invest in its military or accept the status quo.
   (a) If R accepts the status quo, the game stops and payoffs are realized.
   (b) If R invests in her military, p_t shifts to $p_{t-1} + \Delta$ and the game continues.
2. D either makes R an offer X_t such that $f(X_t) = p_t$ or decides to fight a war.
   (a) If D fights a war, the game enters war and payoffs are realized.
   (b) If D makes an offer X_t the game continues.
3. R either allows $Q_t^R | X_t$ to pass and the game repeats, or R chooses to fight a war.

After $\lceil 1/\Delta \rceil$ periods, if the game has not yet stopped, Nature forces R to stop it.

This new sequence of moves highlights three differences between this version of the model and the base-line model. First, R now explicitly chooses to invest in her military and this choice influences the balance of power. Second, competition now takes the form of major war, rather than containment. As we shall soon see, relative power (p_t) influences how states perform in war. Third, D now makes offers that depends on relative power. As a result, D's offer no longer shift by 1 issue across rounds. Rather, the rate of change in D's offer depend on the rate of shifting power Δ conditional on R's military investment. I now explain these differences more precisely.

Result I'll now illistrate that given all of these changes, I can produce equilibrium behavior similar to the behavior described in proposition 2.2.[67] Including shifting power and war creates a number of interesting dynamics that should be explored in a paper-long analysis. It is possible to find conditions that will sustain a variety of different thresholds for T_1^*. Here I solve for the condition where D sets the threshold at $T_1^* = N/4$. As a result, only the greediest

[67] By similar, I mean I can still derive predictions 1-6 based on the equilibrium result.

type under-states her intentions and pools her message with a sub-type $f(\hat{V}) = N/4$. The reason is two fold. First, is a tough case because it requires D to wait the longest not knowing what R will do and make the most concessions. As a result, R grows strong and has the most to gain from war (D has the most to lose). Second, the solution is simple and gets the core intuition across. The example establishes that the strategic behavior I argue for in the manuscript survives as an equilibrium in more complex strategic setting that explicitly rely on shifting power and war to motivate offers.

Proposition C.1 *When:*

$$T^*\Delta < \frac{N - \Delta N - w}{N} \tag{29}$$

$$T^*\Delta < \frac{4(\Delta N + w)}{5N} \tag{30}$$

can be solved for $T^*\Delta < 1/4$ *and:*

$$\frac{\Delta(NT^* - T^* + 1)}{4} < c_w + c_i < \Delta N \tag{31}$$

then the following strategies are an equilibrium.

- *R observes her type,* V*, then:*
  - *If* $f(V) \leq N/4$ *R signals honestly* $\sigma_t(V)$ *in the first round.*
  - *If R is the greedy type* $f(V) = N$ *R signals dishonestly* $\sigma_1(\hat{\omega})$ *in the first round such that* $f(\hat{\omega}_i) = N/4 < f(V)$*.*
- *D processes the signal in one of two ways:*
  - *If the signal implies:* $f(\hat{\omega}_i) < N/4$*, then D adopts beliefs* $\beta_t|\sigma_1(\hat{\omega}) \implies pr(V = \hat{\omega}) = 1$*.*
  - *If the signal implies:* $f(\hat{\omega}_i) = N/4$ *then D adopts beliefs that R's true type is either* $pr(V = \hat{\omega}) = 4/5$ *and* $pr(V = \boldsymbol{1}) = 1/5$*.*

- *D offers elements equal to 1 in $\hat{V}$ in random order.*
- *R types that signaled honestly invest in her military in every period $t \in \{1, T^*\}$ then stops and accepts the status quo.*
- *The greedy type that signaled dishonestly invest in her military in every period.*
- *D fights a war at $T^* + 1$ if R has not stopped the game.*

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

This equilibrium result has all the same features as the baseline result: a threshold emerges at T^* which is the total number of concessions that D prefers to make, rather than fight a war. All types that value fewer than $N/4$ issues signal honestly and reveal their type. Greedy types pool at the threshold with types that value exactly $N/4$ issues. As a result, following a signal that implies R values $N/4$ issues, D is unsure of R's type but is willing to make concessions up until that point. If R invests beyond T^* then D believes that R is greedy and chooses war.

To start, I focus on types of R that pool on a message that implies: $f(\hat{V}) = N/4$. I show that assuming no other types pool on this message, that D and R's strategies are incentive compatible. I then show that types that send a unique separating message that implies: $f(\hat{V}) < N/4$ strictly prefer to do so over pooling with a message $f(\hat{V}) = N/4$.

Starting with types that pool on a message that implies: $f(\hat{V}) = N/4$ I'll show that T^* arises endogenously because of the risk that D is willing to accept given D's shifting beliefs at different points in the game, and given the conditions under which different types of R are willing to invest knowing that investment will trigger conflict. As we shall see, $T^* + 1$ is the point where D prefers to fight if D knows that R is the greediest types and R invests in her military But at that point, only the greediest type wants to invest.

Starting with D's incentives, at period $T^* + 1$ D must prefer to fight a war rather than not fight. However, in periods $t < T^* + 1$, D prefers to make concessions rather than fight. This

works because at $T^* + 1$, if R invests, D updates his beliefs about R's type. On the path, in period $T^* + 1$, D believes that only the greediest types will keep investing. As a result, D prefers war at $T^* + 1$ following R's military investment if: $N(1 - \Delta(T^* + 1)) - w > 0$. The LHS is D's value for fighting a war given R has made $T^* + 1$ investments in her military and R is the greediest type. The RHS, is D's value for not fighting and allowing R to take all the issues. This solves for condition 29 as desired.

Once D is certain that R is greedy and condition 29 is satisfied, D will never delay war. Consider an arbitrary period $T^* + k > T^* + 1$, then D only waits to fight when $N(1 - \Delta(T^* + k)) - c_w > N(1 - \Delta(T^* + 1)) - c_w$ which cannot be satisfied. It follows that if D beliefs that only the greediest types will invest at $T^* + 1$, then D fights conditional on R's investment at that point.

Turning to D's incentives in periods $1 \leq t \leq t^*$. In the first period, D prefers to wait if $N(1 - \Delta) - c_w < N(1 - \Delta t^*)(1 - \lambda_g) + ((1 - \Delta t - \Delta)N - c_w)\lambda_g$. The LHS is D's value for fighting in the first period. The RHS has two components. The first is D's expected utility if R invests T^* times then stops multiplied by the probability that R is not the greediest type $(1 - \lambda_g)$. The second term is D's value for fighting at $T^* + 1$ assuming that R keeps investing and exposes herself as the greediest type. Per the argument in the baseline model, $\lambda_g = 1/5$ given that the only types with $N/4$ core interests pool with the greediest types. On the path, D's beliefs about R's type are constant from $t \in \{1, T^*\}$. It follows that if D is wiling to make concessions in the first period, D will keep making concessions in periods that his beliefs are constant.[68] Since D values all issues equally, D is indifferent between offering strategies. Thus, D is willing to play the offering strategy that offers equal parts of all of $\hat{V}$'s stated core interests.

Turning to R's incentives it must be that the greediest types want to invest at $T^* + 1$ and face war, but no others do. The greediest types incentives at $T^* + 1$ are characterized by: $\Delta(t + 1)N - c_w - c_i(t + 1) > \Delta tN - c_i t$. The LHS is what the greediest type gets for

[68] A similar argument can be made if R signals a type that values fewer core interests than $N/4$.

investing and fighting. The RHS is what the greediest type gets for accepting the status quo. This solves for the right-most inequality in condition 31. When this is satisfied, the greediest types will invest at $T^* + 1$ and face war.

I argue that types that hold $N/4$ core interests pool with greedy types until $T^* + 1$ then they stop investing and separate. For the equilibrium to be separating, types that value $N/4$ core interests must prefer to accept the status quo rather than fight at $T^* + 1$. Under the assumption that $T^*\Delta < 1/4$, then these types prefer not to fight if: $\Delta T^* - c_i T^* > \frac{N}{4}\Delta(T^*+1) - c_i(T^*+1) - c_w$. The LHS is this type's value for accepting the status quo at t^* and the RHS is this types value for investing one more time and facing war. Notice the LHS is identical to the greediest type's value for accepting the status quo offer at T^*. The reason is that the offer is sufficiently small such that it is concentrated only in the $N/4$ issues that both the greediest type and the signaled type value 1. The RHS of the inequality is different because R's value for winning all the issues is different. R's total value for all the issues is: $\frac{N}{4}$. As a result, the limited aims type gains less from fighting than the greediest type would. This solves for the right-most inequality in 31. When this is satisfied, those that value less than the greedy types separate from the greedier types because they have less to gain from fighting a war.

In periods $t < T^*+1$, types that value $N/4$ core interests have identical preferences to the greedier types because each military investment leads them to receive valuable concessions. It follows that they will invest in all earlier periods if the greedier types will.

Discussion I've shown that given a model where power shifts across periods if R chooses to militarize, and power determines C and D's value for war as well as the bargains they strike, I can produce similar results to the baseline model. By similar, I mean that all the core predictions I outlined in the manuscript appear in both models. Yet there are important differences. First, there are fewer conditions under which war can produce this equilibrium. For R to want to invest in the face of war, power must shift sufficiently fast such that greedy

types prefer to make one more investment rather than live with the status quo war (Δ must be sufficiently high). Similarly, D's cost of war must be sufficiently low so that D prefers to fight at the threshold rather than accept nothing (c_w must be low enough). But if Δ is too high and c_w is too low, that D prefers to fight in the first round rather than give R a chance in the hopes that R is limited. In practice, declining powers often can use both containment and war to thwart a rising power. As a result, we might expect the communication and coordination I predict early in power transitions under a broad set of conditions. However, the type of competition that arise when greedy types reveal their true preferences clearly depend on the rate of shifting power, the cost of war and containment and the effectiveness of different forms of competition.

C.2 R values both core and peripheral issues + period-payoffs and discount factors

In the baseline model I assumed that rising powers only value their core interests. This assumption is plausible because taking and holding territory is expensive (Brooks, 1999). When states expand into new territories, they face enormous set-up costs (both financially and politically) for establishing local institutions, repressing local resistance groups who do not want a change in government, teaching the inhabitants of a new land a new language and so on. The financial reward is often larger from engaging in trade, rather than conquering a foreign land outright even absent foreign resistance.[69]

Nevertheless, in bargaining models we typically assume that rising powers value all issues at least a little bit. This might reflect a situation where states held one dominant principle, but also cared a little bit about a few other principles as well (Gilpin, 1983, makes this argument).

To address this concern, I'll now describe a model where all types of R care about all issues

[69]It is trivial to see that my results will hold if R values or core interests $H > 1$ and all peripheral interests $L < 0$.

positively. They just care about their core interests more than their peripheral interests. I'll show that so long as R cares about core interests much more than peripheral interests, that I can find an equilibrium that produces the exact same 6 core predictions I report in the paper. In fact, the equilibrlium I find greatly expands the conditions under which D can find a threshold T_1^* that will support an equilbirium result that allows me to relax $\mathcal{A}_1$ and still produce an informative equilibrium.

The reason my result strengthens is that R sends a slightly different message where all types pool on a first period message $f\hat{V} = T_1^*$. The major difference is that types who hold less than T_1^* core interests now over-state their demands. But because so many types over-state their demands, D is now even more optomistic that R will stop making demands at T_1^*. This signal can be so effective that it drives D to accept a larger threshold than what D would have accepted in the baseline model with a message $y(\sigma(\hat{V} \leq T_1^*))$. In this way, this message actually increases the amount of information that R can reveal about the scope of his demands because she faces an incentive to coordinate over preference order.

Adjustment to the baseline model In the baseline model I assumed that different types of R had different value matrices In these value matrices, a type valued core interests 1 and peripheral interests 0. Here I replace the 0 entries in R's value matrices with l (for low value) and all 1 entries in R's value matrix with h (for high value). I assume that $h > l > 0$. I define H_i as the count of all of the issues that type ω_i values h. To be clear, I still assume that R's type is drawn from the same underlying distribution of types and D still values all issues 1. All other features of the model are the same.

Result For now, I'll make one additional assumption:

$$\mathcal{A}_3 : f(P) \times l < c.$$

This assumptions implies R does not care about her peripheral interests too much relative

to what she will pay if she faces competition.

Proposition C.2 *Given $\mathcal{A}_3$ holds, if $\frac{T_{\omega_i}}{f(P)+c} \geq \lambda_{T1}$ can be solved for some arbitrary first round signal $\sigma_1(\omega_i)$, and $f(P)l < c < f(P)h$ then a critical threshold T_1^* emerges that implies the following equilibrium strategy exists:*

- *R observes her type, V, then sends a message that implies $f\hat{V} = T_1^*$. This implies:*
  - *If $H_i < T_1^*$ R over-states her aims and signals her super-type that implies $\hat{H}_i = T_1^* > H_i$.*
  - *If $H_i = T_1^*$ R signals honestly.*
  - *If $H_i > T_1^*$ R under-states her aims $\sigma_1(\hat{\omega})$ in the first round such that $\hat{H}_i = T_1^* < H_i$. Further, the signaled type $\hat{\omega}$ is a randomly selected sub-type of the true type such that $\hat{V} \subseteq V$.*
- *D offers elements equal to h in $\hat{V}$ in random order.*
- *All types $H_i \leq T_1^*$ stop the game after T_1^* if D has not. All types $H_i > T_1^*$ do not stop the game.*
- *D stops the game at $T^* + 1$ if R has not.*

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

The equilibrium result and conditions share all the features of proposition 2.2 that I make predictions about in predictions (1-6). As a result, I just explain the differences. A threshold T_1^* arises endogenously. Greedy types send messages at the threshold and accept concessions up until that point. In equilibrium, D's strategic behavior relies on the fact that greedy types and limited aims types separate their behavior following T_1^* periods. If R does not stop the game at $T_1^* + 1$ it implies that R prefers to invest and face competition, rather than reveal that they were dishonest. Only types that have $f(H_{\omega_i}) > T_1^*$ have concessions

they value H that they have not yet received an offer for. These types prefer to fight if $f(P)H > c$. For the equilibrium to hold, it must be that limited aims types prefer to stop the game if they know they will face competition otherwise if: $f(P)L < c$.

The second difference is that all types pool their messages at the threshold. This result is no different for types $f(H_{\omega_i}) \geq T_1^*$. It is different for types with more limited aims: $f(H_{\omega_i}) < T_1^*$. In the baseline model, these types signaled honestly making their limited aims known. Here they pool their signal at the threshold.

Surprisingly, pooling at threshold increases the threshold that D is willing to accept.

Lemma C.3 *Assume that $f(P)L < c < f(P)H$ is satisfied, then if $\frac{T_{\omega_i}}{f(P)+c} \geq \lambda_{T1}$ can be solved for some arbitrary first round signal for a value of $T > 1$, then T_1^* is at least as large and possibly larger for the equilibrium described in the extension.*

The reason is that T_1^* arises endogenously, based on what D's expectation is about R's type given a first round signal λ_{T1}. In particular, λ_{T1} was the probability that R would accept the status quo following T_1^* concessions. When all limited aims types signaled honestly, the probability that R would stop at the threshold was the probability that $f(V) = T_1^*$. But when all types pool at the threshold, the probability that R will stop at the threshold is $f(H_{\omega_i}) \leq T_1^*$. Yet in period period $T_1^* + 1$, if R fails to stop the game, D rules out all types that value $f(H_{\omega_i}) \leq T_1^*$ in both games. Thus, when more limited aims types pool on a message T_1^*, D is more confident that R has limited aims. But once R violates the threshold, D is just as alarmed in both cases.

For example, consider the baseline game with $N = 64$ total issues. Suppose $T^* = 4$. If R sends a message that signals a type at the threshold, D rules out all 64 types with $f(V) = 1$ and believes that R is honest with probability $16/(16 + 4 + 1) = 16/25$. If R does not stop the game in period 5, D updates beliefs that R will stop following 16 concessions with probability 4/5. In the extension, R's first round signal $T^* = 4$ leads D to believe that R will stop following four concessions with probability $(64 + 16)/(64 + 16 + 4 + 1) = 80/85$.

But if R invests in the 5th period, D believes that R will stop following 16 concessions with probability 4/5.

Lemma C.4 *When $\mathcal{A}_3$ is violated, but $\mathcal{A}_1$ still holds, there is a unique equilibrium in which R does not stop the game in the first period and D chooses competition.*

The proof is obvious. R cares so much about her peripheral interests that she is willing to face competition and then try to take them in the face of contianment. As a result, it does not matter what R's type is, she will never stop the game because she cares about all issues a lot. In effect, her core and peripheral interests are so close together that all types are basically greedy.

Discussion This extension illustrates two points. First, even if R values her peripheral interests more than 0, I can still find an informative equilibrium that leads to the same predictions as my baseline model. The message R sends is slightly diffferent (all types pool at the threshold) but the strategic implications are the same. In fact, my results are even stronger in this extension even though less types send a first period honest message. The reason is that some rising powers want to under-state their motives but others want to overstate their motives. These competinging incentives interact such that R can more easily exploit her incentive to coordinate on preference order to reveal information about her scope as well.

It is important to note that this message also survives in my baseline model. In fact, it is the message that produces the strongest possible result for my theory.[70]

Second, the results require that R only care about her peripheral interests a little bit. When R cares about her peripheral interests a lot, it is as if all types are greedy. In this case, D will always compete in the first period. This clarifies an important scope condition for my argument that fits with my idea of principles and matches the historical record. It

[70] As I discuss in the paper, I do not focus on this message because its results are so much stronger than the other informative messages that survive a D1 refinement. I chose to focus on a message that is representative of the informative nature of all the messages that survive D1.

must be that there are large differences between how much R values her core and peripheral interests in a way that drives R to emphasize one principle over all the others. When it is unclear what R's core interests are, then my results fall apart.[71]

C.3 R can become greedy over time

In the baseline model, I assumed that R's preferences were fixed. That is, all rising powers knew their long-term strategic intentions right from the start, and no rising power realized that she wanted more. In practice, power transitions last for decades and rising powers may acquire an appetite for expansion along the way. What happens if R starts off with limited aims but then acquires greedy aims during the power transition?

Adjustments to the base-line To answer this question I adjust the model as follows. 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 rounds where in each round:

1. R sends a message about which elements of Q it values, and then either decides to stop investing in her military or not.
   (a) If R stops investing, the game stops and payoffs are realized.
   (b) If R does not stop investing, the game continues.
2. D transfers a single element to R or decides to stop the game.
   (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 payoffs are realized.
   (c) If no one has stopped the game, Nature re-assigns R's type as $V = \mathbf{1}$ with $1 - \gamma$ probability and keeps R's type the same with γ probability.

[71] This fits somewhat with competition dynamics that surrounded Frnaco-Prussian bargianing circa 1886. The Prussians asked for a place in the sun. The British did not know how to interperet this claim, and thought that the Prussians may hold broad interest on contesting all territories even if they prioritized a few. This led to early mistrust and competition.

3. Nature forces R to stop the game after nxn rounds.

The one change is the addition of the probability γ into the model that R becomes the greediest type. This set-up is a tough test of shifting preferences because it assumes that R can only grow more aggressive, and if R does grow more aggressive, she becomes the most aggressive type. Further, once R is the most aggressive type, there is no going back: R will want to take over the whole world forever.

I assume that R's strategic behavior maximizes R's current preferences and does not condition on an expectation that R's preferences might change. However, I assume that D's behavior takes into account that R's preferences might change along the way. This assumption fits closely with the logic of shifting preferences. At any point in the power transition, the leader of R knows what she wants to achieve and will put in place a strategy to achieve it. Although the country's preferences might change (perhaps through a leadership change), the current leader will prioritize achieving her own goals (which are private). In contrast, declining powers must factor in who they will face in the present and in the future. They want the strategy that will maximize their own preferences given who they face at different stages of the power transition. If R is likely to become aggressive, they want to account for that.

Result It turns out that the result is nearly identical to the baseline game. The only difference between this result and the baseline model is that D's critical threshold is now lower. This follows from two effects of γ and the conditions for separation between greedy types and limited aims types. I've already shown that greedy types play the exact same strategy as limited aims types up until the threshold. This is not different for types that turn greedy. It doesn't matter when R becomes greedy (so long as it is before T^*) because all greedy types play the same strategy as limited aims types until that point. As a result, D is willing to make concessions up until the point where R has revealed that her initial message was dishonest because only at that point does the greedy type separate.

D's strategic choice to stop the game now must factor in the probability that R starts greedy and the probability that R becomes greedy at some point before the threshold. Consider that if D waits T periods before enacting competition, then the probability that R never turns greedy before D stops the game is: γ^{T-1}. As a result, the probability that R turns greedy is $1-\gamma^{T-1}$. D's critical threshold now hinges on:

- a threshold T_{ω_i} as the fewest concessions that will completely satisfy type ω_i.
- a probability $\lambda_{Tj} : pr(r_R = T_{\omega_i} | \beta_t(\sigma), O, h(O, \sigma))$.
- a probability $1-\gamma^T$ that D becomes greedy.

At the beginning of the game, the probability that R starts off wanting more that T_{ω_i} does or becomes greedy is: $1-\lambda_{Tj}+\lambda_{Tj}(1-\gamma^{T_j-1})$. Here the new term $\lambda_{Tj}(1-\gamma^{T_j-1})$ is the probability that R started off with an honest signal and then became greedy somewhere along the way.

Discussion All D cares about is whether or not R will stop at the threshold. D does not care if R never intended to stop, or if R honestly wanted to stop at the beginning and changed her mind a long the way. When I allow for the possibility that R can become greedy into the model, D simply factors that in as additional risk that R will not stop at the threshold. But T^* is a function of the risk that R will reveal herself as greedy at the threshold. Thus, when the risk that R is greedy increases, D simply sets a lower threshold.

This result illustrates an important scope condition for my theory that apply when three conditions hold: (1) rising powers have an unstable government at the onset of a power transitions such that declining powers believe that the rising power's government will be over-turned during the power transitions with high confidence; (2) The declining power believes that any alternative government is much more likely to be much greedier than the existing government; (3) the cost of competition is low. Given these three conditions, declining powers will choose competition in the first period, not because the rising power

is likely to be greedy, but because they are likely to become greedy over time. Of course, at more moderate levels of these conditions, declining powers simply accept less risk by accepting a lower threshold T_1^*.[72]

C.4 D starts out thinking that R is much more likely to be greedy than have limited aims.

In the baseline model, I assumed that R was much more likely to be a limited aims type than the greediest type. However, D may start out more worried that R has greedy aims.

Adjustments to the base-line and results To address this concern, I adjust the distribution from which R's type is drawn: $(\Omega \{\omega_1\})$ so that there is an equal chance that R values 1, 4, 16, ... N issues. For example, in the 8x8 example, I assume there is a 1/4 chance R is the greediest type, and a 1/4 chance R is one of the single-issue types.[73] Otherwise, the game is identical.

The result is similar to the case in Appendix C.3 where R can become greedy as γ shifts over time. When R is more likely to be greedy, D is less willing to test R's motives. This does not automatically drive D to select competition because competition is costly. Instead, D sets a lower threshold T_1^*. Given the adjusted type-space where R is equally likely to value 1 or N issues, I can find an equilibrium that matches proposition 2.2 if I replace $\mathcal{A}_1$ with:$f(P) + c > N/4 + 1$.

Interestingly, there is no longer an upper bound on $\mathcal{A}_1$. The reason is that D can now always credibly promise to switch to competition if R violates the original threshold. Thus, raising the chance that D is greedy makes D less willing to test if R has limited aims, but also makes it more likely that D can credibly promsie to switch to competition if D discovers that R's motives are aggressive.

[72] I can't think of a historical case where these two conditions hold in a way that would ruin my predictions.

[73] I assume an equal probability that R is one of the single issue types—but this turns out also not to matter.

In general, as the liklihood that R is the greediest possible type increases, the risk threshold that D is willing to accept decreases holding constant the cost of competition. If R is certainly the greediest possible type, I cannot produce my equilibrium result. However, given that there is a positive probability that R is one of the single issue types, there always a cost of competition where I can find my equilibrium (it must be that $f(P) + c$ is very close to N).

Discussion This result illustrates another important scope condition for my theory that applies when (1) declining powers are very confident that the rising power is a very greedy type; (2) the cost of competition is low. Given these conditions, declining powers will choose competition in the first period because the risk is too great. Of course, at more moderate levels of these conditions, declining powers simply accept less risk by accepting a lower threshold T_1^*.