Core values, multi-dimensional salience, and trust

September 21, 2023

Abstract

Theorists argue that costly actions are the rationalist solution to trust problems. But in many historical cases, costly actions fail to engender mistrust, and avoiding them fails to engender trust. How should costly actions affect trust and competition? I argue the answer lies in contextual information Defenders collect about how a Challenger's motives connect to the situational characteristics of contested issues. I develop a general method to model this variation where (1) limited-aims Challengers hold core values, and intensely value a few, highly salient core objectives; (2) Defenders can partially observe whether a contested issue is highly salient for the Challenger's core values. The model produces equilibria that past models of trust cannot sustain, but fit puzzling historical episodes. When a contested issue is salient to core values, costly actions do not engender mistrust or competition. When these issues also hold low instrumental salience, costly actions can theoretically engender trust and peace. I resolve persistent empirical puzzles by showing the standard logic mainly applies when costly actions serve peripheral interests. Two case vignettes illustrate my implications over many trust domains. My modeling technology is useful for any multi-issue reputational problem (e.g, resolve).

Word Count: 9,727

Table of Contents

1 Introduction

Trust problems are central to competition during power transitions (Powell 1996), conventional and nuclear arming (Debs and Monteiro 2014; Slantchev 2010; Bas and Coe 2016), arms control (Coe and Vaynman 2019), economic cooperation (Svolik 2006), leadership targeting (McGillivray and Smith 2008), and elite competition (Svolik 2012). The security dilemma (Jervis 1978) that drives the tension in realism, (Waltz 1979; Kertzer and Mcgraw 2012) and the spiral that it can generate, are trust problem (Jervis 1978). Many believe that Sino-American relations (Glaser 2015) and cyber-conflicts (Axelrod and Iliev 2014) are problems of trust as much as resolve. Across these domains, scholars have identified costly signals as the common rationalist solution (eg Kydd 2005; Glaser 2010). When a state militarizes, takes territory, exploits coercive threats, violates treaty commitments, it signals greedy motivations (Yoder 2019; Haynes and Yoder 2020; Coe and Vaynman 2015; Garfinkel and Dafoe 2019). Avoiding these actions reassures.

While the logic is compelling, the evidence is mixed. In many empirical domains, uninformed Defenders scrutinize a Challenger's costly military actions, but their trust does not reliably shift as signaling theory expects. For example, costly signaling theory well explains how the United States reassured the Soviets through arms control agreements at the Cold War's end (Kydd 2005). But it struggles to explain how a few years later the US *engendered* trust across Europe by *expanding* its military footprint to new territories (Kydd 2001). During power transitions declining powers are surprisingly hopeful as rising power rapidly militarize and instigate crises (Schweller 2004; Goddard 2018). Notably, British perceptions of Hitler's motives did not change for years as Hitler engaged in rapid militarization and territorial conquest (Wark 1985). Yarhi-Milo (2014, pp85-88) documents an especially puzzling case where Halifax *raised* his trust in Hitler after Hitler re-militarized the Rhineland. Some use this pattern to question if costly signaling theory is sufficiently reliable for drawing predictions (Rosato 2007; Mearsheimer 2001).[1] Others use it to motivate bureaucratic (Schub 2023) and psychological (Yarhi-Milo 2014) explanations. As I discuss more in the conclusion, the Senate Select Committee on Intelligence has used the disconnect to chastise the intelligence community for failing to warn of China's vast aims following repeated crises and militarization efforts (Schiff 2020).

How should states interpret costly signals? I argue the answer lies in information Defenders collect about how a Challenger's interests connect to the situational characteristics of contested issues. I develop a general method to model this variation, and embed it into a simple model of trust. Like past models I focus

[1] These cases do not fit offensive realist predictions either because we often observe no updating.

on uncertainty over the Challenger's intrinsic motivations that vary from limited aims to greedy (e.g., Kydd 2005). I assume only greedy Challengers are sensitive to instrumental value that facilitates hegemony, profit or expansion for its own sake (Little and Zeitzoff 2017). I then make three empirically plausible changes.

First, I model a second dimension (Trager 2015) of intrinsic motivations that I call the *core values* dimension. These are values that even limited aims Challengers care a lot about. Substantively, this could represent a limited state's concern for security from foreign threat (Glaser 1995). But consistent with how certain constructivists (O'Neill 1999), liberal theorists (CARTER, ABRAMSON, and YING 2022; Abramson and Carter 2016) and comparativists (Mylonas 2013) discuss intense but limited motivations, the core values dimension could also represent limited aims Challengers that intensely value ethnic or religious nationalism, or unifying previously held borders under one government (Goemans and Schultz 2017; Schultz 2017). Second, I embrace the fact that different contested issues are salient to the Challenger for different reasons. Thus, a *core interest* is an issue that is highly salient for the Challenger's core values. For example, Taiwan is a core interest of China's because even if China was purely motivated by restoring its position in Asia (and had no hegemonic intrinsic motivations), then China would care intensely about Taiwan.[2]

Finally, and most importantly, I assume that Defenders can estimate the reason that a contested issue is salient for the Challenger. Past scholars argue (Press 2007; Goddard 2018) or take as assumption (Sartori 2005) that Defenders cannot observe issue-specific variation. But in real life Defenders research the attributes of contested issues (Lowenthal 2019). They know when an issue is well endowed with natural resources, is populated by the Challenger's ethnic or religious kin, or whether the issue is useful for the Challenger's security. They may not trust the Challenger because they do not know if the Challenger holds greedy or limited intrinsic motivations. But they can more reliably estimate if the Challenger's costly actions serve the Challenger's core vales, or if an issue is salient for instrumental reasons.

The Defender's knowledge of issue salience allows the Defender to draw contextual inferences from events with different kinds of salience. This illuminates surprising predictions about the relationship between military actions, trust, and patterns of competition that cannot be sustained in standard signaling models (see Ramsay 2017). First, when a contested issue is a core interest and also rich in instrumental wealth, then the Challenger's decision to rapidly militarize, violate arms control agreements and even invade will not engender mistrust or trust, and cannot effect the Defender's choice to compete. Second, when a contested issue is especially salient for the Challenger's core values, but offers little instrumental value, then

[2]The reason is Taiwan is highly salient to China's core values.

the Challenger's costly military actions engender trust. The reason is that limited aims Challengers are sensitive enough to core value that they are willing to bare significant material costs to prevail. In contrast, greedy types prefer to wait for a more lucrative opportunity. By the same logic, avoiding costly military actions can engender mistrust and competition. This means costly military actions over core interests should not signal greedy intentions, and avoiding costly actions over core interests should not signal limited aims. Rather, the standard logic only applies when Challengers contest issues that are weakly salient for the Challenger's core values but salient for instrumental reasons.

Overall, these results show that we cannot draw simple linear predictions that expanding alliance commitments, rapid militarization, offensive arming, or territorial demands communicate aggressive intentions. But I also show that context does not prevent states from drawing reliable inferences from past actions (Press 2007) because a rival's historical context is somewhat observable. Defenders who are uncertain of a rival's true intentions can still use their historical knowledge to adjudicate how close an issue is to that rival's core interests.

I develop one extension that explains why the classic logic rarely appears in commonly studied power transition cases. In it, the Challenger can strategically select which issue to contest (Trager 2015). Challengers with limited aims always select their core interests first. Therefore, the Defenders cannot learn from early costly actions. Other extensions show that my results are robust to continuous variation in the Challenger's motives (Spaniel and Smith 2015), and different kinds of situation-specific uncertainty commonly studied in theories of resolve (Press 2007).

As I present my theoretical results, I re-visit the puzzling cases discussed above. I show that many well-fit my logic of costly signaling once I code them for salience. In section 3, I report two cases vignettes that illustrate my general logic plausibly applies across many empirical domains, and further illuminates the most surprising theoretical results. In one case, the US decision to avoid a costly humanitarian intervention against Rwanda engendered mistrust amongst those concerned the US may exploit human rights as a pretext. In another, Russia's military intervention in Georgia raised the trust of certain influential US policymakers.

By embedding historical context into a model of trust, I answer Fearon and Wendt (2002)'s call to forge a closer connection between those that identify the determinants of foreign policy orientations—including research in comparative, and identity politics (O'Neill 1999; Finnemore 1996; Trager and Vavreck 2011; Jackson and Morelli 2011; Mylonas 2013, beyond those above)—with strategic theories of trust. This holds important implications for patterns of competition (Kydd 2005; Glaser 2010), and threat perceptions (Jervis

2010; Yarhi-Milo 2014; Friedman 2019) we expect from rationalist states. My procedure for modeling motives and salience is easily integrated into models of other multi-issue reputational problems, such as resolve (Dafoe, Renshon, and Huth 2014; Schultz and Goemans 2019), war outcomes (Mastro and Siegel 2023; Jordan 2021), alliance credibility (Benson and Smith 2022), diplomacy (Malis and Smith 2019; Lindsey 2017; Trager 2015), and border coordination (CARTER et al. 2022). I discuss these and policy implications for Sino-American relations in the conclusion.

2 Model

First I set-up the model. Second, I explain how my assumptions depart from existing modeling assumptions, and motivate my novel parameters with a discussion of US perceptions of China. Motivating the parameters after presenting the model is recommended when the parameterization is based on realistic variation (Paine and Tyson 2020; Morrow and Sun 2020). Third, I characterise the Defender's trust problem. Fourth, I report my theoretical results. Finally, I consider modeling adaptations.

2.1 Set-up

I study a strategic interaction between a Challenger C and a Defender D. The basic structure of the action space is common to theories of international reputation over repeated interactions.[3] Table 1 presents the sequencing. These steps are designed to generate a classic trust problem for D (and a reassurance problem for C). At step (1), Nature privately assigns the Challenger's strategic intentions as greedy C_g or limited C_l such that $pr(C = C_l) = \alpha, pr(C = C_g) = 1 - \alpha$. At steps (2) and (4) Nature determines the situational features of a contested issue, then C is given an opportunity to take a costly military action (or not) that will raise the chance of a concession (or not). This could represent $C's$ military intervention, regional military deployments, election meddling, etc. For simplicity, I discuss this as a revision opportunity.

At step (3) D is given the opportunity to engage in competition or not. If D engages in competition, D locks in a payoff for the second revision opportunity at $1 - p - w > 0$. If D does not select competition, D's payoff at step (4) depends on C's action.[4] I notate D's strategy as $s^D(a), a \in \{c, nc\}$. I notate C's

[3] See Dafoe et al. (2014) for general discussion. The basis is the finite-period chain-store model (Kreps and Wilson 1982). Closer adaptations to international trust and resolve include Gurantz and Hirsch (2017); Yoder (2019). We make the simplifications proposed by Goldfien, Joseph, and McManus (2022) who also study more nuanced preference functions.

[4] Competition could represent war, containment, sanctions, or D's choice to withdraw from institutions that would allow C to exploit her future.

type-specific strategy as $s^g(a_1, a_2), s^l(a_1, a_2)$ where $a_1 \in \{r_1, nr_1\}$, and $a_2 \in \{r_2, nr_2\}$.

My innovation comes in how I structure (a) the Challenger's intrinsic motivations; (b) the situation-specific features of the revision opportunity; and (c) the information the Defender holds. I summarizes these features in Table 2, which correspond with Nature's motives in Figure 1. Table 3 describes payoffs given all the possible strategy profiles.

I assume that issues C can contest at (2) carry two dimensions of value that I label *instrumental* and *core*. These are represented by indexes i, n (Because C is taken) respectively. Thus, at 2(a), Nature draws the amount of instrumental $\theta_i \sim f_i()$, and core $\theta_n \sim f_n()$ salience of the issue in dispute. Where both distributions are supported on non-negative real numbers. These draws are revealed publicly.

At the beginning of the game, when Nature privately assigns C's type, Nature also privately assigns C's sensitivity to instrumental and core value. $C_g's$ sensitivity parameters are π_i, π_n, and $C_l's$ are ρ_i, ρ_n. As Table 3 shows, the value the Challenger accrues from revision is sensitivity $\times$ salience. For example, if C selects revision in the first period, C_g adds $\theta_i\pi_i + \theta_n\pi_n - k$ to his utility. C_l adds $\theta_i\rho_i + \theta_n\rho_n - k$ and D accrues 0. If C does not select revision, C accrues 0 and D accrues 1.

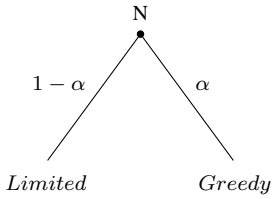
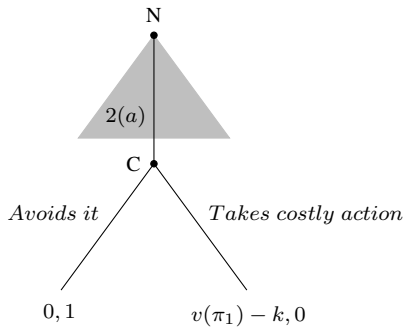
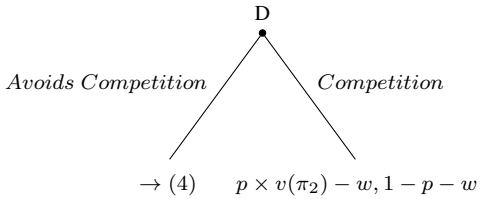
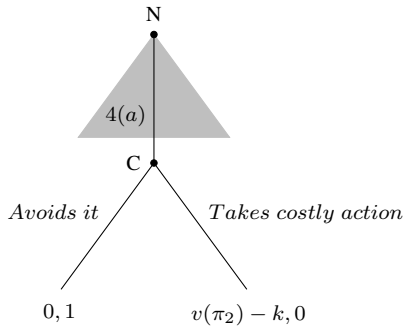
I make the following assumptions to meet my understanding of limited and greedy Challengers:

- Both types are sensitive (ie., accrue utility from) instrumental and core value: $\pi_i, \pi_n, \rho_i, \rho_n > 0$.
- C_g is more sensitive to instrumental value than C_l: $\pi_i > \rho_i$.
- C_g is more sensitive than C_l overall: $\pi_i + \pi_n > \rho_i + \rho_n$
- C_l is more sensitive to core values than instrumental values: $\rho_n > \rho_i$

The second period is terminal. It's main point is to create a trust/reassurance problem. Thus, I simplify here (which does not substantively alter my results). I assume that the greedy type always values the second period issue $H > k$. The limited type values the issue H with probability λ, and 0 with probability $1 - \lambda$. If there is no competition and C selects revision, C_g adds $H - k$, and C_l adds either $-k$, or $H - k$ depending on λ. D adds 0. If there is no competition and C does not select revision, C accrues 0 and D accrues 1. Competition forces second period competition payoffs. D gets $1 - p - w$, C_g gets $Hp - w$ and C_l gets $\lambda Hp - w$. All utilities accrue at the end.

My primary outcome of interest is trust. Consistent with the existing literature, I define trust as the Defender's beliefs that the Challenger holds limited (not greedy) intentions. I define $\alpha_0 = \alpha$ as the Defender's initial (prior) belief that the Challenger's aims are limited. I am interested in how the Challenger's

Table 1: Sequence of moves

Step (1): Challenger's (C) Disposition	
	Nature determines if the Challenger has greedy or limited aims (private for C).
Step (2): C's opportunity to signal intentions through costly action.	
	Nature determines C's situation specific cost/benefits. Then, C decides to take a costly military action (or not) to capitalize on that opportunity/not. Revision entails $v(\pi_1)$ benefits and k costs. These vary with the situation and C's type. $v(\pi_1)$ is a placeholder that we specify in Table 3.
Step (3): Defender's (D) Opportunity for Preventive Competition	
	D decides to intervene and deny the Challenger from further opportunities for revision (leading to fixed competition payoffs) or not (leading to a second revision opportunity).
Step (4): C's revision opportunity.	
	Nature determines C's situation specific variables. Then, C decides to capitalize on that opportunity/not. Revision entails $v(\pi_2)$ benefits and k_2 costs. These vary with the situation and C's type. Critically, greedy types have higher expected values here, which implies revision is more likely.

Note: Below I specify the information structure and payoffs at steps 1, 2(a), 4(a)

Step	Description
1	Nature makes C limited with pr. α and assigns sensitivity (or intrinsic motivation) parameters ρ_i, ρ_n. Nature makes C greedy with pr. $1-\alpha$ and assigns sensitivity parameters π_i, π_n. C's intrinsic motives are private.
2(a)	Nature determines issue-specific core θ_n and instrumental θ_i salience. The issue-specific salience is public.
4(a)	Nature determines if the limited aims Challenger values the second revision opportunity High (with probability λ) or not high $(1-\lambda)$. I assume low λ, guarantees limited types rarely select second-period revision.

Note: Steps refer to Figure 1. Note: these are simplest assumptions in main model. Extensions relax many of them.

Table 2: Information and issue-specific variation

no	Strategy	U^g	U^l	U^D
1	$s^C(nr_1, nr_2), s^D(nc)$	0	0	2
2	$s^C(nr_1, r_2), s^D(nc)$	$H-k$	$H-k \vert \lambda=1, -k \vert \lambda=0$	1
3	$s^C(r_1, r_2), s^D(nc)$	$\theta_1\pi_i + \theta_n\pi_n - k + H - k$	$\theta_1\rho_i + \theta_n\rho_n - k + \lambda(H-k)$	0
4	$s^C(r_1, nr_2), s^D(nc)$	$\theta_1\pi_i + \theta_n\pi_n - k$	$\theta_1\rho_i + \theta_n\rho_n - k$	1
5	$s^C(r_1, .), s^D(c)$	$\theta_1\pi_i + \theta_n\pi_n - k + pH - w$	$\theta_1\rho_i + \theta_n\rho_n - k + \lambda pH - w$	$1-p-w$
6	$s^C(nr_1, .), s^D(c)$	$pH - w$	$\lambda pH - w$	$2-p-w$

The strategy notation for Challengers is r_t, nr_t meaning revision / not in period t. For Defenders, c, nc means competition / no competition. For presentational ease, no. 3 is written as an expectation given λ. Strictly it should be presented conditional on λ.

Table 3: Utilities

first period choice effects trust. I generally define the Defender's posterior belief as α_1. But this belief will depend on equilibrium strategy. I define $\alpha_1|r_1$ as the Defender's level of trust if he observes the Challenger revise these status quo, and $\alpha_1|nr_1$ the Defender's level of trust if the Challenger avoids revision. My secondary outcome is competition/peace. As I explain soon, and is standard, I focus on parameters wherein competition and trust are directly related.

2.2 Motivating the parameters

Others study complex variation in state motives but cannot produce my findings. Many include situational variation and dispositional uncertainty (see Debs 2020, for review). But they only study one dimension of preferences (ideal point variation Schultz and Goemans 2019; Spaniel and Smith 2015, is similar). Thus, C_l cannot intensely value specific issues. Some model multi-dimensional motives (Trager 2015; Battaglini 2002; Penn, Patty, and Gailmard 2011; Joseph 2021). But they aggregate dispositional and situational factors because they assume C's preference for each issue are private, and drawn independently from a common distribution. Thus, D cannot infer C's preference for one dimension by learning $C's$ value for the other. To

my knowledge no study disaggregates situational and dispositional characteristics of the same parameters[5] and allows the Defender to observe the former. I show that a specific combination of these factors unlocks novel insights. In what follows, I substantiate my assumptions with a discussion of China.

In recent years, China presented the United States with a trust problem (Glaser 2015). The United States believed that "Inferring China's *strategic intentions* was the single most difficult and important challenge that we faced."[6] Consistent with trust theory, the United States believed that it was plausible China held limited intentions. But China never claimed, and the US never believed, that if China's aims were limited, China was motivated exclusively by security. Rather, China reliably articulated a set of core values that China claimed drove its intense but limited objectives. In early years, China articulated core values through diplomatic encounters with the United States. More recently, China wrote them in Defense White Papers. For example, in 2000 it wrote, "China has always attached primary importance to safeguarding state sovereignty, unity, territorial integrity and security, and has been working hard for a peaceful international and a favorable peripheral environment for China's socialist modernization drive."

Of course, the United States knew that China held an incentive to misrepresent. However, China's expression of core values impacted US estimates as far back as Kissinger's secret visit to China. Shortly After Kissinger's visit, in November 1970, the CIA produced National Intelligence Estimate (NIE) 30-7-70 titled "Communist China's International Posture." In the second paragraph, it explains that China's "basic goals" are most likely determined by *one of two* intrinsic motivations. Either China sees itself as "a great power and leader of the world revolution (C_g) or as a more traditional but highly nationalistic country *concerned primarily* with Asian interests (C_l)." The NIE then spends two pages detailing China's vision of the Middle Kingdom to explain exactly what it means for China to to be sensitive to nationalistic value. Interestingly, NIE 30-7-70 asserts that these two potential basic goals will partly drive China to contest *different* issues. This is consistent with the view that revolutionary communist states disclaim their ethnic and religious identity to some degree, and instead prioritize global values (Fawn 2003). By contrast, states whose basic goals are nationalism, intensely value nationalistic goals.

Overall, my parameterization of intrinsic motives captures this variation. Like NIE 30-7-70, I assume the US is uncertain about which set of intrinsic motivations China holds (α). But the US also believed that if China held limited aims, it was sensitive to its stated core values (ρ_n is high),[7] but less sensitive to

[5] i.e, others study situational costs and dispositional values.

[6] Author's interview with CIA Director for Analysis, Mark Lowenthal (2014).

[7] The standard assumption that limited aims Challengers want all issues less (Kydd 2005) cannot account for this possibility.

instrumental value necessary to pursue world revolution (ρ_i is low). If China was greedy, it was much more sensitive to instrumental value (π_n is high) and also likely somewhat sensitive to core values that China has asserted (ρ_n is also high).

To be clear, this framework for parsing different dimensions of value extends beyond Sino-American relations. A key ingredient in any modern national security strategy is listing core values that are "broadly conceptual and do not substantially change except over the long term; they are the slowly evolving essence of a nation's character".[8] But in each case, the articulated core values will be different. For example, Britain understood Prussian core values as the unification of German-speaking territories. The Monroe Doctrine articulated a different set of core values for the United States (Gilderhus 2006). While each set of core values is different, the common feature is that these expressions articulate what C values intensely if its aims are truly limited.[9]

Turning to the issue-specific parameters. I make the non-controversial assumption that different contests can be valuable to the Challenger for different reasons. This means that Tibet is highly salient to China for historical and nationalist reasons (θ_n is very high), and less salient for instrumental reasons (θ_i is lower). By contrast, China has no historical attachments to the oil fields of Saudi Arabia, and therefore θ_n is low, and θ_i is high.

I also make the more controversial assumption that D can reasonably estimate the different sources of value for many particular contests (cf Press 2007; Rosato 2015).[10] To be clear, my position is not unique among empirical scholars. In the context of resolve, Kertzer (2016) argues that dispositional information is hard to observe but situational information is easier. But he does not explore the strategic implications, nor does he apply it to trust problems.

My assumption is intuitively true in the examples above. If I can code Tibet's historical relevance to China, and Saudi Arabia's vast oil fields along distinct dimensions of salience, it is likely US intelligence analysts can also. Declassified evidence shows that the CIA codes salience dimensions in this way during important contests. For example, as the 1959 Taiwan Straits Crisis was unfolding, SNIE 100-4-59 clarified that Taiwan was highly salient for national reasons (θ_n is high) but less salient for economic productivity, or regional expansion (θ_i is low). Of course, Taiwan's economic salience has changed. Shifting salience does

[8] https://www.atlanticcouncil.org/content-series/strategy-consortium/elements-of-national-security-strategy/.

[9] For a strategic explanation for how this can arise, see (Trager 2015; Battaglini 2002).

[10] Press (2007) argues ambiguity is pervasive, then argues it makes signaling resolve (not trust) difficult. Goddard (2018) argues that contests hold no objective value that can be estimated in advance. Rather their interests form through rhetoric.

not impact my theory because reputation follows from how states behave given the value they accrue from contesting an issue at a specific moment (Dafoe et al. 2014). Therefore, the most important feature is the value they accrue at that moment. The key feature for my theory is that the Defender can and does estimate θ_i, θ_n at the time that Challenger takes a costly military action. SNIE 100-4-59 confirms this for the 1959 Taiwan Strait Crisis the CIA could confidently estimate along which dimension this contest was salient to China at that moment.

What is more, the United States well understands how China's core values interact with these situation-specific features to generate core interests. I asked Director of National Intelligence Blair how the intelligence community went about estimating the issues that China wants if China held limited intentions. Without prompting, he used the exact language of core interests. In his words, "determining core interests has a long and honorable place in analysis, [however] that's not that difficult. You ask any 50 China-analysts and they'd give you that same list of things that I came up with." Consistent with my theory, he argued these core interests represent what China would want if China held limited aims based on his understand of China's expressed core values. Then analysts use their knowledge of issue salience to code China's core interests.

In certain cases, the truth likely lies somewhere in between the DNI's strong confidence and Rosato (2015)'s deep uncertainty. The Defender likely estimates with some confidence how salient a specific issue is for the Challenger on these different dimensions. In section 2.5.4, I discuss an extension that models this and other forms of situation-specific uncertainty.

My bet is that this D can exploit his knowledge of θ_i, θ_n to draw nuanced inferences about whether the Challenger's motives are greedy or limited based on the Challenger's costly actions. In short, the CIA knows facts about Taiwan and Saudi Arabia. Therefore they use those facts to draw different inferences about China's strategic aims if China takes (avoids) a costly military action to contest Tibet or Saudi Arabia. China likely knows this, and likely strategically alters its choices.[11] While this point appears obvious, existing models assume situation-specific information is private to C and drawn anew each contest (eg Sartori 2005; Yoder 2019), and therefore do not account for it systematically (see Ramsay 2017, for review). While informal theories discuss it, they ultimately simplify it out when they examine the strategic dynamics in their theories (Glaser 2010). In the analysis that follows, I show that how D infers trust in response to the Challenger's costly military actions will hinge on whether D knows the contested issue is highly salient

[11] Mastro (2022a) argues that failing to parse intentions, objectives and information has caused much confusion over China.

along either θ_i, θ_n or both dimensions.

2.3 Preliminary Analysis: Generating the Defender's Trust Problem

For C's costly military actions to reassure, D must face a trust problem at the moment D decides between competition and peace. For a trust problem to arise, D's competition choice must hinge on perceptions of C's intrinsic motives. If D is sufficiently trusting, D selects peace, hoping that C will avoid revision in the second period. If D is insufficiently trusting, D locks in the competition payoff. Define a **trust threshold:**

$$\alpha^* = \frac{1-p-w}{1-\lambda}$$

This gives us the following result.

Lemma 2.1 ***Bounding parameters to guarantee a trust problem.*** *If*

$$\mathcal{A}_1 = H > k$$

$$\mathcal{A}_2 = 1 \geq \frac{1-p-w}{1-\lambda} \geq 0$$

hold then the following strategies are on path in every Perfect Bayesian equilibrium. At the second revision opportunity, $s^g(r_2), s^l(r_2|\lambda = 1, nr_2|\lambda = 0)$. *D's equilibrium decision to compete is* $s^A(c|\alpha_1 < \alpha^*, nc|\alpha_1 > \alpha^*)$.

$\mathcal{A}_1$ ensures that the C selects second-period revision for high-valued issues. The greedy type selects revision more frequently because limited aims types are less likely to locate high valued revision opportunities ($\lambda < 1$). This brings us to D's strategy. D wants to avoid competition if he believes C is unlikely to select second-period revision. Given C's strategy, D's competition choice is determined by:

$$1 - p - w > \alpha_1(1-\lambda)$$

This re-arranges to our trust threshold (α^*). When $\alpha_1 < \alpha^*$ D strictly prefers competition to peace.

$\mathcal{A}_2$ simply states that the trust threshold must be in $0, 1$. This is necessary to generate a trust *problem*. If $\alpha^* > 1$, or $\alpha^* < 0$ D would select competition or peace respectively no matter how much he trusted C.

For the remainder of the analysis I assume $\mathcal{A}_1, \mathcal{A}_2$ and solve for perfect Bayesian equilibrium (PBE).

2.4 Analysis

When the Defender faces a trust problem, the limited aims Challenger faces a reassurance problem. If the limited aims Challenger can credibly reveal her motives, then the Defender selects peace. If she cannot, the Defender selects competition. In the model, the first revision opportunity gives C a chance to communicate her intentions through her costly military actions. The standard logic of costly signaling is that C engender trust (mistrust) if she avoids (takes) costly military actions. But, as stated in the introduction, the evidence for this pattern is mixed.

My goal is to understand how variation in issue-specific salience moderates the logic of costly signaling. With this goal in mind, I summarize all the equilibria that can emerge in Table 4. I plot these equilibria as a function of the instrumental and normative salience of the contested issue in Figure 1.[12] In what follows, I clarify how salience moderates the logic of costly signaling with reference to these results.

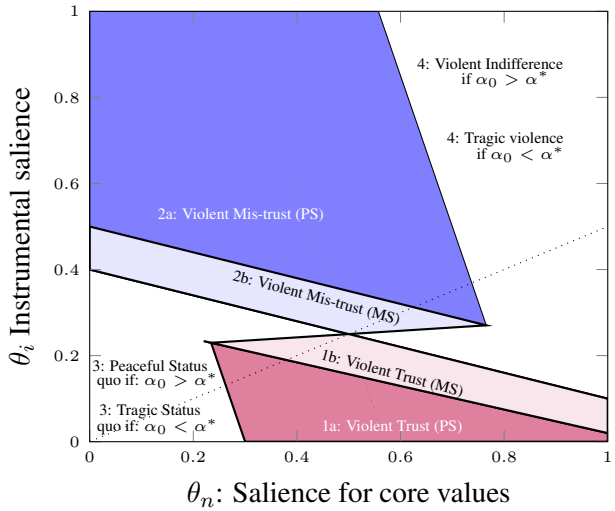
[12] Strictly, an equilibrium includes a strategy profile for every $f_i(\theta_i)$, $f_n(\theta_n)$. But since these are drawn publicly before the first action, we can effectively treat the θ_i, θ_n as prior parameters.

Table 4: Equilibrium descriptions

Equilibrium	Necessary Prior	C's first period action: revision (r_1), not (nr_1)	D's reply (Competition/peace) given trust	Violence affects Trust: $\alpha_1\|r_1 - \alpha_1\|nr_1$
1a. Violent Trust (ps).	NA	C^l: revision, C^g: no revision	Competition if no revision, peace otherwise.	1
1b. Violent Trust (ms)	$\alpha_0 < \alpha^*$	C^l: revision, C^g: mixes	Mixes if revision. Competition otherwise.	$1 - \alpha^*$
1b. Violent Trust (ms)	$\alpha^* < \alpha_0$	C^l: mixes, C^g: no revision.	Mixes if no revision. Peace otherwise.	α^*
2a. Violent Mistrust (ps)	NA	C^l: no revision, C^g: revision	Competition if revision, peace otherwise.	-1
2b. Violent mistrust (ms)	$\alpha_0 < \alpha^*$	C^l: no revision, C^g: mixes	Mixes if no revision. Competition otherwise.	$-\alpha^*$
2b. Violent mistrust (ms)	$\alpha^* < \alpha_0$	C^l: mixes, C^g: revision.	Mixes if revision. Peace otherwise.	$-(1 - \alpha^*)$
3. Tragic status quo	$\alpha_0 < \alpha^*$	All: no revision	Competition	0
3. Peaceful status quo	$\alpha^* < \alpha_0$	All: no revision	Peace	0
4. Tragedy of politics	$\alpha_0 < \alpha^*$	All: revision	Competition	0
4.Violent indifference	$\alpha^* < \alpha_0$	All: revision	Peace	0

Eq. Numbers correspond with Figure 1. ps = pure strategy, ms = mixed. Positive updating in col. 5 means D is this amount more trusting after observing revision than if D observed no revision. 'mixes' in col. 4 implies competition and peace are both possible.

Figure 1: Equilibria as a function of the salience of a contested issues.



Notes: Assumes $k = .4, \lambda = .1, H = .6, w = .2, \pi_i = 1, \pi_n = .6, \rho_i = .2, \rho_n = .7$. Above the *dotted line* C_g values the contested issue more than C_l. In the unshaded regions there is a unique pure strategy equilibrium that survives the intuitive criterion (conditional on priors). In the *shaded regions* costly military signals support trust-building equilibria where competition is contingent on revision choices. Dark shades are pure strategies, light shades are mixed strategies. Violent trust are red, violent mistrust are blue. We bound $\theta < 1$ for visual ease.

First, I provide a general result that explains when the standard predictions cannot apply.

Result 1 If the contested is salient enough along the core dimension ($\theta_n > \frac{\lambda(H-p)+k(1-\lambda)+w-\theta_i\rho_i}{\rho_n}$) then C's choice to take costly military actions cannot decrease trust or cause competition; and C's choice to avoid costly military actions cannot raise trust, or prevent competition.

See Lemma A.1. For learning to occur, Challengers with different motives must play different strategies. But C_l is sensitive to core values. Thus, if an issue is salient enough to C's core values, then the limited

aims Challenger is sensitive enough that C_l prefers revision no matter the consequences. D appreciates this problem. Because D can estimate the salience of a contested issue to $C's$ core values, he knows when the C_l holds direct incentives to take costly military actions. In these cases, D expects C_l to incur the costs of revision. Thus, D does not infer that C_g is more likely to take costly actions over these issues.

We see this on the right hand side of Figure 1. Notice that there is no equilibrium that matches the classic logic of costly signaling. To be clear, result 1 is not an equilibrium claim. What is more, it does not imply that costly signaling generates the standard results when the stated condition is violated. We can see this in the plot. Even when θ_n is reasonably low, we can arrive at some nuanced results that the standard logic finds hard to explain.

In what follows, I characterize the implications of equilibrium behaviors. There are many equilibria, including mixed strategy equilibria, pooling equilibria, and pure strategy separating equilibria. However, they can all be categorized into three types: costly signals engender trust (violent trust) and reduce the risk of competition, costly signals engender mistrust (violent mistrust) and raise the risk of competition, and costly signals hold no affect (violent indifference). Fortunately given my interests, these results are neatly organized as functions of the salience parameters θ_i, θ_n. Thus, I present the results informally that capture how the salience of the contested issue effects trust and competition across the categories of equilibria. These informal results are carefully worded to capture differences between mixed strategy and pure strategy, and carefully worded so as not to contradict nuances that arise in modeling extensions discussed in section 2.5. The analysis for all equilibria appears in Appendix A.2.

I start with the classic result.

Result 2: The conventional trust prediction, violent mistrust, bears out if and only if the Challenger contests an issue that is highly salient for instrumental values, but not salient for core values (θ_i is high and θ_n is low). In this case costly military actions engender mistrust and raise the chance of competition. Avoiding costly actions engenders trust and peace.

The mechanism is described by Kydd (2007). C_l cares little about purely instrumental issues relative to C_g. When a Challenger fight for these issues, they signal that they care a lot about instrumental value, and, therefore, are likely greedy. The more interesting component of this result are the salience dimensions over which it applies. Returning to Figure 1, we can think about the classic trust model as representing the case where $\theta_n = 0$. In this case, only the instrumental parameters matter (θ_i, π_i, ρ_i), and therefore we can arrange preferences in the standard way from greedy to limited. In this case we can only achieve the classic

results. If we accept that preferences are multi-dimensional (θ_n gets larger), this classic result is harder to substantiate in equilibrium. When θ_n is high enough we cannot substantiate it at all.

Result 3 helps clarify the dis-connect between existing theoretical expectations that follow when preferences are arrayed on one dimension and historical patterns. Past theories provide important insights for contests that C_l values low. Specifically, those issues and territories that are not salient to $C's$ core values. But they should not apply for the issues that fill our history books such as the Rhineland Crisis, the Taiwan Straits Crisis, Stalin's demands for Poland, etc.

We now turn to the novel results.

Result 3 If D can estimate with moderate confidence that a contested issue is highly salient to C's core values and also for instrumental reasons (θ_i, θ_n are both high), then $C's$ costly military action minimally alters trust $\alpha_0 - \alpha_1 \approx 0$. Because $D's$ trust does not change, $D's$ competition choice does not change:

- If $D's$ initial trust is high ($\alpha_0 > \alpha^*$) C always selects revision, and peace persists because D cannot parse greedy or limited Challengers.
- If initial trust is low ($\alpha_0 < \alpha^*$) C always selects revision, D selects competition because he cannot parse greedy or limited Challengers.

When both kinds of salience are high then both the greedy and limited aims Challenger care intensely about the contested issue. It follows that both take costly military actions. If $D's$ beliefs do not vary, then $C's$ costly military action cannot influence $D's$ choice to compete. This leads to a novel result: when D can observe that issues are highly salient along both dimensions, and initial trust is high, we observe indifference to C's costly military actions. C selects revision, D is not alarmed by C's violent action, and does not select competition.

Result 3 explains many of the motivating examples. For example, Hitler's decision to renounce the League, rapidly militarize, and to orchestrate the Anschluss all served issues that scored high on both kinds of salience. Existing accounts assert that the rational British reaction was mistrust. They then explain Britain's failure to warn as wishful thinking, select attention, shifting discount factors, or other psychological, beuracratic, or normative mechanisms (see Gilbert 1972; Barnett 1986; Goddard 2018; Yarhi-Milo 2014; Edelstein 2019; Weisiger 2013).[13] While there is some micro-evidence to support these accounts, my theory suggests that perfectly rational British elites would have arrived at the same estimate. For example, because they knew the salience of Austria, they realized that even if Hitler held limited aims he would annex

[13] Ripsman and Levy (2008) argue that British competition choices were based on expectations of shifting power. This is important, but it does not explain why threat perceptions did not change (Wark 1985).

it.

Result 3 also clarifies an unresolved debate between offensive and defensive realists. Offensive realists characterize a tragedy of politics wherein limited Challengers covet territory to serve there limited (usually security) motivations. But in doing so they trigger additional mistrust and competition (Mearsheimer 2001). Defensive realists argue that this argument makes strong assumptions about prior beliefs. They suggest that if Defenders start out more trusting, then the Challenger can avoid revision to signal their limited aims (Kydd 1997). My model reveals that neither argument is quite right. In a situation where an issue is highly salient to instrumental and core values, it is not that case that violence triggers additional mistrust. Rather, trust is invariant. If trust starts below the threshold $\alpha_0 < \alpha^*$, then the Defender selects competition. But the Defender was always going to do that. Thus, a costly military action does not trigger competition in a case like this. Rather, competition was inevitable to begin with. It is also not the case, as defensive realists contend, that more trusting priors opens up costly signaling opportunities under these conditions. The reason is that C_l values these issues intensely. Thus, C_l will not forgo revision to engender trust.

Result 4: If D can estimate with moderate confidence that a contested issue is highly salient to C's core values, but also holds low instrumental salience (θ_i low, θ_n is high) then the Challenger

- (a) engenders trust if she takes a costly military action, raising the chance of peace
- (b) engenders mistrust if she avoids a costly military actions, raising the chance of competition.

Results 4 is the opposite of the conventional wisdom from defensive realists and other trust scholars (Kydd 2005). Both Challenger-types are sensitive to the costs of military action (k). Whether these costs are worth the gains of military actions depends on how much utility C extracts from shifting the status quo. Challengers with different motives are sensitive to different sources of value. Thus, issues with low instrumental value differently effect C_l and C_g. This change alters their direct and indirect incentives to take costly military actions. C_l does not care much about instrumental value, and so variation along θ_i does not influence her direct incentives. The contested issue still scores high on core values. Therefore the C_l is willing to pay a large situation-specific cost to contest the issue. Both types also face indirect incentives. $C_l's$ incentive is heightened because she rarely cares intensely about issues. C_l knows similar opportunity are unlikely to arise in the future. This amplifies $C_l's$ incentives to act now.

C_g also cares about core values. However, C_g is especially sensitive to instrumental value. Thus, the reduction in instrumental value diminishes $C_g's$ direct incentive. When the situation-specific cost of revision is sufficiently large, the cost-benefit proposition is not worth it for C_g. What is more, C_g knows that she is

likely to value future issues high. As a result, the greedy type can wait because a more lucrative opportunity is likely to present itself.

A careful reader may note that the type-specific sensitivity parameters mean that extreme cases arise where $C_l's$ total utility for revision exceeds $C_g's$.[14] Is this condition necessary to achieve violent trust? The answer is no. In Figure 1, the dotted line plots the point where both Challenger-types accrue the same utility from revision. Above the line, C_g accrues more utility. Even in this case, the violent trust equilibrium persists. The reason is that even though C_l has weaker direct incentives to participate in violent trust, she holds stronger indirect incentives because it is unlikely she will care about future issues.[15] As a result, it is possible to sustain violent trust even when the Challenger faces a revision opportunity that the greedy type values strictly more.

Result 4 explains the most puzzling cases discussed in the introduction. For example, Kydd (2001) argues that costly military actions surrounding NATO expansion raised trust in the post-Cold War period. Notably, the US accrued little instrumental gain from this expansion. NATO did not provide additional rents, and it exposed the US to more commitments it was concerned to make, and the US gained little additional persuasion in Europe. If the US was motivated by expanding globally, this would not be an efficient use of resources. Different still, Yarhi-Milo (2014) asserts that the Rhineland crisis facilitated trust for Halifax. The Rhineland had enormous nationalist value to Germany. But the instrumental value of *re-militarization* is less clear. After all Germany collected taxes from the Rhineland and benefited from economic production pre-crisis. Thus, the additional instrumental value of military control is less clear.

Overall, my theory also explains other puzzling features of these case. For example, Yarhi-Milo (2014) notes that Defenders often focus intensely on specific cases (e.g, pp45-46). She argues that this intense focus on *litmus tests* is not rational because the contests are not especially costly. She concludes that they must follow from a sub-optimal obsession with vivid images. My theory provides a rationalist answer: Defenders seek out distinctive contests along salience dimensions θ_i, θ_n that create unique opportunities for costly signaling.

[14] To some degree this fits the NIE-30-7-70 discussion wherein revolutionary communist states disclaim nationalistic and religious identity.

[15] Adding in other forms of costly signaling allows for violent trust under even broader sets of preferences. For example, if states lose military power taking a contested territory.

2.5 Adapting the model.

I develop four modeling extensions. The purpose is to verify my interesting results can survive given plausible complications. Thus, I confine my analyses to these points. However, several extension illuminate how modeling issue salience can drive novel empirical and theoretical insights. Thus, the short extensions illustrate the framework's value for future theoretical research.

2.5.1 Strategic selection

Like others, I assume that Nature determines the situation-specific features of contests (Trager 2015; Sartori 2005; Yoder 2019). This is plausible in certain trust scenarios. In the 1990s the US did not choose which countries would initiate human rights atrocities. Thus, it did not select into opportunities where it could pursue humanitarian intervention or not. During the 1995 Taiwan Straits Crisis, a poorly timed visit to the US, and Taiwanese rhetoric arguably provoked a scenario for China to respond to. But in other cases, especially power transitions, Challengers can usually select contests. For example, Hitler chose the order of expansion to a large degree. How does strategic selection effect my result?

In Appendix A.3, I assume C selects between contesting one of two issues, or no revision. I focus on the interesting case where the C_g prefers issue 2 over 1, and C_l prefers issue 1 over 2. Strategic selection reveals something surprising under conditions that fit initial periods of power transitions. When there are outstanding issues that are salient to the Challenger's core values, then limited aims Challengers choose the contest with the highest salience to them. Thus, we arrive at a variant of result 1: costly military actions do not engender mistrust because the limited aims Challenger always selects the most salient contest.

Does this drive costly signals engender indifference, or trust? The key factor is how much C_g accrues from his favourite issue relative to C_l's favourite issue. When C_g's value from both issues is close, C_g pools with C_l to avoid competition (result 3). But if C_g vastly prefers his favourite issue, C_g selects a contest that C_l would never have selected. This triggers mistrust and competition. The classic costly mistrust logic follows when k exceeds C_l's value for all available issues.

Consider the implications for the Anglo-German case. The British could not discern Hitler's intentions because Hitler kept selecting issues that were salient for his core nationalist values. Hitler's actions and statements leading up to Munich clearly show his selection process understands the strategic game my theory illuminates. Consistent with my logic, a week before taking the Sudetenland Hitler publicly declared, 'I am

asking neither that Germany be allowed to oppress three and a half million Frenchmen, nor am I asking that three and a half million Englishmen be placed at our mercy. Rather I am simply demanding that the oppression of three and a half million Germans in Czechoslovakia cease and that the inalienable right to self-determination takes its place.' But this claim was not genuine. One year earlier, Hitler called a secret meeting of Nazi leadership where he detailed his true foreign policy vision. No minutes were taken. However, attendees later wrote down their recollection of Hitler's desires in the Hossbach Memorandum (which they kept secret). Hitler explained his desire to take Eastern Europe and exterminate the Slavic populations, then expand across Western Europe (Weinberg 1980, pp39-40). Why keep these specific interests secret but publicise a desire to take the Sudetenland? Consistent with my theory, Nazi elites understood that British elites knew the territories salient in Germanic values, and would not increase mistrust against these contests. Debate between Hitler and other leading Nazi officials further shows that they knew the invasion of Eastern Europe would would trigger 'major war' with Britain because Eastern Europe was not historically salient for Germany (Hillgruber 1974, pp5-22). As a result, the Nazis first selected the contests that were salient for Germany's asserted core values, even though Hitler coveted Eastern Europe more for its instrumental value. Because they selected these issues first, British elites were unable to discern Germany's intentions for a long time.

2.5.2 Continuous intrinsic motives

Appendix A.4 studies a model where the Challenger's motives vary along a continuum from extreme limited to extreme greedy (Spaniel and Smith 2015). Doing so leads to substantively similar results. However, the continuity of types precludes complete separating equilibria. Instead, a cut-point on the Challenger's type emerges. In the violent trust equilibrium, more limited Challengers select revision, and greedier challengers avoid revision. This allows for only partial (rather than complete) trust-building. But the direction of trust is still moderated by salience as my theory expects.

2.5.3 Situation-specific uncertainty

Appendix A.5 studies a model where D is uncertain about $C's$ situation-specific cost of revision (k becomes k_1, k_2) (Press 2007). I show my interesting predictions are robust. Consistent with Sartori (2005), I find that uncertainty moderates what D infers, but the direction of the inference remains constant.

2.5.4 Ambiguity over core interests

Appendix A.6 studies a case where China knows whether Taiwan is truly salient for China's core values or not (i.e, C knows if θ_n is high or low), but the US is uncertain, and privately develops an intelligence estimate of Taiwan's core salience (D gets a private signal of θ_n that is accurate with probability ψ, where ψ represents the quality of intelligence).

In the context of the reputation for resolve, Press (2007) argued that an inability to determine situational features, including the salience of a particular issue, would cause signaling to degrade. Rosato (2015); Goddard (2018) asserts the concern extends to classic trust theory. I explore it explicitly. First, I consider the affects under the standard conditions. If we start in a world that produces the classic costly signaling argument (i.e, result 2), then ambiguities over issue salience can diminish signaling opportunities. However, the scope of uncertainty must be severe. To have any affect at all, the range of different estimates on θ_n must be so large as to shift C_l's preferences from competition to peace. Even in this case, signaling is still possible, but tragic mistrust occurs when D substantially under-estimates θ_n. For a signaling equilibrium that relies on D's knowledge of issue-salience to degenerate entirely, D must employ low quality intelligence analysts.

I then consider how these ambiguities impact my novel results (result 4). I find that even with moderate quality intelligence analysts that my results survive. I provide additional contextual information to clarify the parameter ranges fit many real cases. In the condition I analyze, ambiguity does not impact the inference of trust following a costly action. The reason is that C_g always avoids revision. Thus, C_l decision to contest an issue over-rides ambiguities. However, ambiguity does attenuate (but not ruin) D's inference that C is greedy following nr_1. We find another surprising result. By taking a costly military action, C_l teach D about the true value of the contested issue. Thus, costly actions can inform about $C's$ intrinsic motives, and the state of the world θ_n. These results illuminate the value of dis-aggregating salience parameters. They show that Press's valuable substantive insights lead to far more nuanced signaling dynamics in the context of trust than he may have envisioned.

3 Vignettes

Throughout the theoretical analysis, I showed how my novel results explain puzzling historical episodes. In what follows I provide two modern vignettes. To be clear, I accept that these cases are multi-causal,

and involve many dynamics beyond rationalist trust. My goal is not to rule out alternative theories and dynamics. Rather, I use vignettes to illustrate that my logic can connect together many different kinds of trust problems in many different policy domains within a come framework of rationalist trust. These vignettes also illuminate some of the most interesting theoretical results. In one case a costly military action engendered trust and cooperation, in the other avoiding a costly military action caused distrust.

3.1 Avoiding military action engenders mistrust: Rwanda and the Clinton Doctrine.

During president Clinton's tenure, the values of human rights increasingly crept into documents that described US values. For example, the section of the 1994 National Security Strategy that characterized US core values included a sentence "Our commitment to freedom, equality and human dignity continues to serve as a beacon of hope to peoples around the world (pii)." As the concern for global human rights grew, "humanitarian intervention became an important pillar in the emerging new world (Lyon and Dolan 2007)." President Clinton publicly supported humanitarian interventions because he believed it would make the US safer (an instrumental reason), and because he believed that protecting human rights globally was intrinsically valuable (Klar 1999).

For many governments in the developing world the ascension of human rights as an important US value caused a trust problem. There was an "overriding concern that states would use the pretext of humanitarian intervention to wage wars for ulterior motives (Goodman 2006)." This trust problem arose at a difficult time because the US was largely unchecked (Kydd 2001). The challenge for the US was in convincing others that it was sincerely motivated by humanitarian concerns, and that it was not using this as a pretext for global expansion during the unipolar moment (Monteiro 2014). In terms of my model, the parameters sub-script n represent human rights as a core value. Thus, if US claims were genuine, then ρ_n was high and ρ_i was low. If the skeptics were right, then π_i operated and was much larger than ρ_i.

This case is especially interesting because crises arose largely at random, and Clinton could chose whether to respond or not. The question is, how did Clinton's costly military actions effect the perceptions of skeptics? I argue it depends on the salience of each crisis for both instrumental reasons and humanitarian reasons. Consistent with my theory, the US interventions we observed had no effect on trust because, as Mason and Wheeler (1996, p100) notes, "the cases which can plausibly be regarded as examples of humanitarian international involve mixed motives." This statement clearly fits my result 3. Mason argues that θ_i and θ_n were both high in all the cases that the US intervened. As a result, those that were concerned about

humanitarian values as a pretext did not draw inferences from these cases because the US would have taken these interventions if its values were genuinely limited, or not.

Clinton's decision not to intervene in key cases also fits my theory. According to standard costly signaling logic the choice to avoid a military intervention should engender trust (Kydd 2005). By contrast, my theory predicts that avoiding a costly military actions can signal mistrust if the issue is highly salient along the core value dimension (θ_n is large) and is not salient for instrumental reasons (θ_i is low). This dynamic closely fits reactions to Clinton's decision to avoid intervention against the Rwandan Genocide. The humanitarian costs of inaction in Rwanda was enormous (θ_n high), but the intervention would not produce economic benefits or strategic value for the United States (θ_i low). Consistent with my theory, "the failure of the West to respond to the Rwandan genocide in 1994 shows that humanitarian claims must compete with other interests (Finnemore 2003, p73)." That is the failure to take a costly military action amplified the skeptics. The reason is that if the US was genuine, this was a key case that they would take up.

3.2 Trust in Russia following the invasion of Georgia.

Since the dissolution of the Soviet Union, the regions of Abkhazia and South Ossetia along the Russo-Georgian border were hotly contested. Russia asserted these regions were salient in ways that other parts of the former Soviet Union were not because they conferred security, identity and prestige (Tsygankov and Tarver-Wahlquist 2009). Thus, even a limited aims Russia would hold high value for them. During the 1990s, Russian values lay dormant because the region was de facto controlled by Russian back separatists. De-factor control meant that the value of further control was low (θ_n was lower). Tensions began to flare after the 2003 Rose Revolution. A pro-Western Georgian government came to power, and Georgia began to more frequently raise Abkhazia and South Ossetia as important issues. What is more, NATO discussed expansion to Georgia, which would cover Abkhazia and South Ossetia. As the fear of losing this core interests increased, so too did the salience that would come from a military contest (θ_n increased) (Tsygankov and Tarver-Wahlquist 2009).

In 2008, Russian-backed separatists in Abkhazia and South Ossetia began shelling Georgia. Conflict broke out. Russia invaded to support separatists. Russia won, and established military bases in these territories. While outraged by Russia's sovereignty violation, many policy analysts believed that Russian actions communicated that Russia was sincere in its limited strategic objective to ensure its leadership position in it's near abroad (Charap 2021).

Recognizing this salience, many western observes did not raise threat perceptions. More surprisingly for the standard costly signaling story, a group of prominent analysts argued that Russia's military intervention was reassuring (result 4). A 2007 announcement that Georgia would eventually join NATO played a major role in this estimate. As O'Hanlon (2021), a member of the State Department's International Security Advisory Board in 2008 explains, "Given Georgia's distance from Europe and the North Atlantic, it was increasingly hard for many Russians to view NATO's interest in Georgian membership as anything more than imperial overstretch, and at their own country's expense (p19)." Consistent with my theory, O'Hanlon concludes that Russia's decision to invade Georgia was the result of "pressures that had been building in Russian minds for many years (p22)" over threats to Russian core interests. As Charap (2021, pp87-92) documents, O'Hanlon was not alone in his conclusion that Russia's costly actions communicated limited aims. Götz (2015) has similarly argued that later Russian deployments to Georgia evidence how "Russia has a genuine national interest in preventing outside powers from acquiring a foothold."

To be clear, many analysts that raised their trust in Russia did so because Russian forces pushed beyond South Ossetia into Gori and other regions that did not clearly fit within Russia's near abroad, before pulling back Charap (2021). They reasoned that the pull-back evidenced Russia's limited aims because Russia invaded less than what it could have. This further supports my theory. It demonstrates that Russia understood that military actions for different territories would engender different reactions. A plausible reason is that Russia's salience claim was lower in these other regions that were historically Georgian controlled.

Also to be clear, not everyone believed that the Georgian invasion signaled Russia's limited aims. However, those that perceived trust had influence. In 2009, the EU issued a controversial report[16] that found "no way to assign overall responsibility for the conflict to one side alone (p327)." Referring to the security context and NATO expansion, the report found that even though Russia fired the first shot, the military conflict "was only the culminating point of a long period of increasing tensions, provocations and incidents" (p326). In 2010, the NATO-Russian Council had their most productive meeting in years, ending in a joint-statement that opens with "We... affirmed that we have embarked on a new stage of cooperation towards a true strategic partnership." Also in 2010, NATO delayed considering Georgian and Ukrainian membership. These events are inconsistent with the predictions of the spiral model (Kydd 2005). In the spiral model, analysts that observe a violent action forge mistrust and respond with increased arming. In this case, several analysts perceived the Georgian invasion as confirming Russia held limited aims, and the Western allies

[16] https://www.nato.int/cps/en/natohq/news_68871.htm

reduced NATO expansion, among other policies. I can explain them through the logic where costly military actions over salient territories can engender trust. What is more, the reactions from NATO illustrates that the implications of trust can be substantial enough to break the spiral of mistrust.

4 Conclusion

Scholars from a broad range of traditions believe that trust is essential for avoiding competition (Weisiger 2013; Glaser 2010; Kydd 2005; CARTER et al. 2022; Yarhi-Milo 2014; Svolik 2006). The standard logic is that costly military actions engenders mistrust and avoiding them reassures (Jervis 1978). Theorists extend this logic to many types of actions and empirical domains in comparative and international politics (Debs and Monteiro 2014; Slantchev 2010; Bas and Coe 2016; Kertzer and Mcgraw 2012; Coe and Vaynman 2019; Svolik 2006, 2012; McGillivray and Smith 2008; Haynes and Yoder 2020; Yoder 2019; Coe and Vaynman 2015; Garfinkel and Dafoe 2019; Trager 2011). But the evidence is mixed, and this causes many to question whether states act rationally in these critical moments (Schub 2023; Schweller 2004; Yarhi-Milo 2014), or if the theory is useful for reliably predicting behavior in any real-life trust setting (Rosato 2007; Mearsheimer 2001).

I argue that contested issues are salient to Challengers for different reasons, and that these salience dimensions are partly observable to Defenders. I develop a general way to model this insight. I embed it into a general formal model of trust-building. This yields systematic predictions about costly military actions, trust and competition that substantially depart from the standard arguments, and reconcile many puzzling cases across different empirical domains with costly signaling theory. Most critically, empirically scholars often focus on high-stakes issues that are highly salient to the Challenger's core values. I show these are the issues where the standard logic should not apply. If an issues is salient for core values and instrumental reasons, then costly military actions cannot effect trust or competition. If an issue is only salient to core values (and not instrumentally salient), then costly military actions can raise trust, and even foster cooperation. Avoiding costly actions can even engender trust. The standard logic only applies if contests are weakly salient for the Challenger's core values.

This vindicates costly signaling theory by explaining it is not supposed to apply in commonly studied cases. Throughout the theory section I showed that the most puzzling great power cases, including British indifference to Hitler's rampage across Europe (Yarhi-Milo 2014), or trust that followed from NATO expan-

sion (Kydd 2001), are all systematically explained once we code them for salience. I used two case vignettes to show that the logic applied to a wide range of trust scenarios.

The results hold broad theoretical implications. Scholars cannot draw simple linear predictions that rapid militarization, offensive arming, or the scope of demands communicates states are aggressive. Rather the historical and cultural context has far reaching strategic implications that can reverses the direction of these common rationalist predictions. My theory provides an overarching framework to systematically explain the conditions under which we should expect updating in each direction.

The results hold critical policy implications. In 2020, the US Select Committee reviewed all estimates of China's intentions. Culminating in a report, "The China Deep Dive: A Report on the Intelligence Community's Capabilities and Competencies with Respect to the People's Republic of China." The report correctly notes that China engaged in military build ups, instigated crises in Taiwan, coerced concessions from Tibet among other costly military actions. It then chastised the intelligence community for failing to warn about China's strategic intentions given these costly military actions. In other documents the Committee members argue this represents an intelligence failure (Schiff 2020). They, and others (Medeiros 2005; Medeiros and Blanchette 2021) call for major intelligence reform. But intelligence estimates are easily explained because all the contested issues are highly salient for China's core values. Thus, my results caution against reforms that will punish analysts for taking into account China's historical and cultural context. This may lead to worse estimates in the future.

My framework for modeling motives is easily implemented into other theories of reputation with multiple contested issues. Consider its likely theoretical and empirical implications for different kinds of resolve (see Dafoe et al. 2014). Some empirical scholars code issue-specific variation (Hensel and Mitchell 2007),[17] and others code state or leader-specific cost-sensitivities (Kertzer 2016). But these researchers do not connect a state's specific core values at a particular moment in history to this issue-specific variation. For example, Goemans and Schultz (2017) show that ethnic salience can impact conflict, but Altman and Lee (2022) show it does not always matter. If we accept that states are motivated by different core values at different moments in history we can likely resolve these discrepancies. On the theory side, scholars model situational and dispositional differences, but do not fully dis-aggregate them to match the rich empirical findings we have (Debs 2020). A key simplification is that states are entirely uncertain about situation-specific

[17] ICEWs also introduces concerns surrounding international law. My theory could potentially amplify its insights by considering when core values and lawful actions intersect.

features (eg Trager 2015; Joseph 2021). It is plausible that avoiding force over a peripheral interest could communicate high resolve to fight for core values in the future. Therefore it is possible to construct settings where the US can avoid a crisis in Europe to communicate an interest in Asia (Hopf 1994); or the choice to renege on one kind of alliance commitment could resolve the commitment problems for others (Mastro 2022b). These kinds of insights will grow in importance as the US must balance its reputation as resolve against multiple threats under budget constraints (Tarar 2023).

References

Abramson, Scott F. and David B. Carter, 2016. The historical origins of territorial disputes, *American Political Science Review*, **110**, 675–698.

Altman, Dan and Melissa M Lee, 2022. Why territorial disputes escalate: The causes of conquest attempts since 1945, *International Studies Quarterly*, **66**.

Axelrod, Robert and Rumen Iliev, 2014. Timing of cyber conflict, *Proceedings of the National Academy of Sciences*, **111**, 1298–1303.

Barnett, Correlli., 1986. *The collapse of British power*, Humanities Press International.

Barnett, Michael, 1999. Culture, strategy and foreign policy change:, *European Journal of International Relations*, **5**, 5–36.

Bas, Muhammet A and Andrew J Coe, 2016. A dynamic theory of nuclear proliferation and preventive war, *International Organization*, **70**, 655.

Battaglini, Marco, 2002. Multiple referrals and multidimensional cheap talk, *Econometrica*, **70**, 1379–1401.

Benson, Brett V. and Bradley C. Smith, 2022. Commitment problems in alliance formation, *American Journal of Political Science*.

CARTER, DAVID B, SCOTT F ABRAMSON, and LUWEI YING, 2022. Historical border changes, state building, and contemporary trust in europe, *American Political Science Review*, **116**, 875–895.

Charap, Samuel, 2021. *Russia's Military Interventions: Patterns, Drivers, and Signposts*, RAND Corporation PP - Santa Monica, CA.

Coe, Andrew and Jane Vaynman, 2019. Why arms control is so rare, *American Political Science Review*, pp. 1–14.

Coe, Andrew J. and Jane Vaynman, 2015. Collusion and the nuclear nonproliferation regime, *The Journal of Politics*, **77**, 983–997.

Dafoe, Allan, Jonathan Renshon, and Paul Huth, 2014. Reputation and status as motives for war, *Annual Review of Political Science*, **17**, 371–393.

Debs, Alexandre, 2020. Mutual optimism and war, and the strategic tensions of the july crisis, *American Journal of Political Science*, p. ajps.12569.

Debs, Alexandre and Nuno P. Monteiro, 2014. Known unknowns: Power shifts, uncertainty, and war, *International Organization*, **68**, 1–31.

Edelstein, David M., 2019. *Over the horizon : time, uncertainty, and the rise of great powers*, Cornell University Press.

Fawn, Rick, 2003. Ideology and national identity in post-communist foreign policies, *Journal of Communist Studies and Transition Politics*, **19**, 1–41.

Fearon, James D. and Alexander Wendt, 2002. *Rationalism v. Constructivism: A Skeptical View*, pp. 52–72, SAGE Publications Ltd.

Finnemore, Martha, 1996. *National Interests in International Society*, Cornell University Press.

Finnemore, Martha, 2003. *Changing Norms of Humanitarian Intervention*, pp. 52–84, Cornell University Press.

Friedman, Jeffrey A., 2019. *War and chance : assessing uncertainty in international politics*, Oxford University Press.

Garfinkel, Ben and Allan Dafoe, 2019. How does the offense-defense balance scale?, *Journal of Strategic Studies*, **42**, 736–763.

Gilbert, Martin, 1972. *The Roots of Appeasement*, Weidenfeld and Nicolson.

Gilderhus, Mark T., 2006. The monroe doctrine: Meanings and implications, *Presidential Studies Quarterly*, **36**, 5–16.

Glaser, Charles L., 1995. Realists as optimists: Cooperation as self-help, *International Security*, **19**, 50–90.

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

Glaser, Charles L., 2015. A u.s.-china grand bargain? the hard choice between military competition and accommodation, *International Security*, **39**, 49–90.

Goddard, Stacie E., 2018. *When right makes might : rising powers and world order*, Cornell University Press.

Goemans, Hein E. and Kenneth A. Schultz, 2017. The politics of territorial claims: A geospatial approach applied to africa, *International Organization*, **71**, 31–64.

Goldfien, Michael A., Michael F. Joseph, and Roseanne W. McManus, 2022. The domestic sources of international reputation, *American Political Science Review*, pp. 1–20.

Goodman, Ryan, 2006. Humanitarian intervention and pretexts for war, *American Journal of International Law*, **100**, 107–141.

Gurantz, Ron and Alexander V. Hirsch, 2017. Fear, appeasement, and the effectiveness of deterrence, *The Journal of Politics*, **79**, 1041–1056.

Götz, Elias, 2015. It's geopolitics, stupid: explaining russia's ukraine policy, *Global Affairs*, **1**, 3–10.

Haynes, Kyle and Brandon Yoder, 2020. Offsetting uncertainty: Reassurance with two-sided incomplete information, *American Journal of Political Science*, **64**, 38–51.

Hensel, Paul R and Sara M Mitchell, 2007. The issue correlates of war (icow) project issue data set: Territorial claims data.

Hillgruber, Andreas, 1974. England's place in hitler's plans for world dominion, *Journal of Contemporary History*, **9**, 5–22.

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

Jackson, Matthew O and Massimo Morelli, 2011. The reasons for wars: An updated survey, *The Handbook on the Political Economy of War*, pp. 34–57.

Jervis, Robert, 1978. Cooperation under the security dilemma, *World Politics*, **30**, 167–214.

Jervis, R, 2010. *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War*, Cornell University Press.

Jordan, Richard, 2021. Symbolic victories and strategic risk, *Journal of Peace Research*, **58**, 973–985.

Joseph, Michael F., 2021. A little bit of cheap talk is a dangerous thing: States can communicate intentions persuasively and raise the risk of war, *The Journal of Politics*, **83**, 166–81.

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

Kertzer, Joshua D. and Kathleen M. Mcgraw, 2012. Folk realism: Testing the microfoundations of realism in ordinary citizens, *International Studies Quarterly*, **56**, 245–258.

Klar, Michael, 1999. The clinton doctrine, *The Nation*.

Kreps, David M and Robert Wilson, 1982. Reputation and imperfect information, *Journal of Economic Theory*, **27**, 253–279.

Kydd, Andrew, 1997. Sheep in sheep's clothing: Why security seekers do not fight each other, *Security Studies*, **7**, 114–155.

Kydd, Andrew, 2001. Trust building, trust breaking: The dilemma of nato enlargement, *International Organization*, **55**, 801–828.

Kydd, Andrew H., 2005. *Trust and Mistrust in International Relations*, Princeton University Press.

Lindsey, David, 2017. Diplomacy through agents, *International Studies Quarterly*, **61**, 544–556.

Little, Andrew T. and Thomas Zeitzoff, 2017. A bargaining theory of conflict with evolutionary preferences, *International Organization*, **71**, 523–557.

Lowenthal, Mark M., 2019. *Intelligence : from secrets to policy*.

Lyon, Alynna J. and Chris J. Dolan, 2007. American humanitarian intervention: Toward a theory of coevolution, *Foreign Policy Analysis*, **3**, 46–78.

Malis, Matt and Alastair Smith, 2019. A global game of diplomacy, *Journal of Theoretical Politics*, **31**, 480–506.

Mason, Andrew and Nick Wheeler, 1996. *Realist Objections to Humanitarian Intervention*, pp. 94–110, Palgrave Macmillan UK.

Mastro, Oriana Skylar, 2022a. Understanding the challenge of china's rise: Fixing conceptual confusion about intentions, *Journal of Chinese Political Science*, **27**, 585–600.

Mastro, Oriana Skylar, 2022b. Reassurance and deterrence in asia, *Security Studies*, **31**, 743–750.

Mastro, Oriana Skylar and David A Siegel, 2023. Talking to the enemy: Explaining the emergence of peace talks in interstate war, *Journal of Theoretical Politics*, **35**, 182–203.

McGillivray, Fiona and Alastair Smith, 2008. *Punishing the prince : a theory of interstate relations, political institutions, and leader change*, Princeton University Press.

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

Medeiros, Evan S., 2005. Strategic hedging and the future of asia-pacific stability, *The Washington Quarterly*, **29**.

Medeiros, Evan S. and Jude Blanchette, 2021. Beyond colossus or collapse: Five myths driving american debates about china, *War on the Rocks*.

Monteiro, Nuno P., 2014. *Theory of Unipolar Politics*, Cambridge University Press, quot;Since the collapse of the Soviet Union, the United States enjoys unparalleled military power. The international system is therefore unipolar. A quarter century later, however, we still possess no theory of unipolarity. Theory of Unipolar Politics provides one. Dr. Nuno P. Monteiro answers three of the most important questions about the workings of a unipolar world. Is it durable? Is it peaceful? What is the best grand strategy a unipolar power such as the contemporary United States can implement? In our nuclear world, the power preponderance of the United States is potentially durable but likely to produce frequent conflict. Furthermore, in order to maintain its power preponderance, the United States must remain militarily engaged in the world and accommodate the economic growth of its major competitors, namely, China. This strategy, however, will lead Washington to wage war frequently. In sum, military power preponderance brings significant benefits but is not an unalloyed goodquot;–.

Morrow, James D. and Jessica S. Sun, 2020. *Models of Interstate Conflict*, pp. 261–276, SAGE Publications Ltd.

Mylonas, Harris, 2013. *The politics of Nation Building*, Cambridge University Press.

O'Hanlon, Michael, 2021. *Beyond NATO*, Brookings Institution.

O'Neill, Barry, 1999. *Honor, Symbols, and War*, University of Michigan Press.

Paine, Jack and Scott A. Tyson, 2020. *Uses and Abuses of Formal Models in Political Science*, pp. 188–202, SAGE Publications Ltd.

Penn, Elizabeth Maggie, John W. Patty, and Sean Gailmard, 2011. Manipulation and single-peakedness: A general result, *American Journal of Political Science*, **55**, 436–449.

Powell, Robert, 1996. Uncertainty, shifting power, and appeasement, *The American Political Science Review*, **90**, 749–764.

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

Ramsay, Kristopher W., 2017. Information, uncertainty, and war, *Annual Review of Political Science*, **20**, 505–527.

Ripsman, Norrin M. and Jack S. Levy, 2008. Wishful thinking or buying time? the logic of british appeasement in the 1930s, *International Security*, **33**, 148–181.

Rosato, Sebastian, 2007. The flawed logic of democratic peace theory, *The American Political Science Review*, **97**, 585–602.

Rosato, Sebastian, 2015. The inscrutable intentions of great powers, *International Security*, **39**, 48–88.

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

Schiff, Adam, 2020. The u.s. intelligence community is not prepared for the china threat, *Foreign Affairs*.

Schub, Robert, 2023. Informing the leader: Bureaucracies and international crises, *American Political Science Review*.

Schultz, Kenneth A., 2017. Mapping interstate territorial conflict, *Journal of Conflict Resolution*, **61**, 1565–1590.

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

Schweller, Randall L., 2004. Unanswered threats: A neoclassical realist theory of underbalancing, *International Security*, **29**, 159–201.

Slantchev, Branislav L., 2010. Feigning weakness, *International Organization*, **64**, 357–388.

Spaniel, William and Bradley C. Smith, 2015. Sanctions, uncertainty, and leader tenure, *International Studies Quarterly*, **59**.

Svolik, Milan, 2006. Lies, defection, and the pattern of international cooperation, *American Journal of Political Science*, **50**, 909–925.

Svolik, Milan W., 2012. *The Politics of Authoritarian Rule*, Cambridge University Press.

Tarar, Ahmer, 2023. Crisis bargaining in the shadow of third-party opportunism, *International Studies Quarterly*, **67**.

Trager, Robert F., 2010. Diplomatic calculus in anarchy: How communication matters, *American Political Science Review*, **104**, 347–368.

Trager, Robert F., 2011. Multidimensional diplomacy, *International Organization*, **65**, 469–506.

Trager, Robert F., 2015. Diplomatic signaling among multiple states, *The Journal of Politics*, **77**, 635–647.

Trager, Robert F. and Lynn Vavreck, 2011. The political costs of crisis bargaining: Presidential rhetoric and the role of party, *American Journal of Political Science*, **55**, 526–545.

Tsygankov, Andrei P. and Matthew Tarver-Wahlquist, 2009. Duelling honors: Power, identity and the russia-georgia divide, *Foreign Policy Analysis*, **5**, 307–326.

Waltz, Kenneth N., 1979. *Theory of international politics*, McGraw-Hill.

Wark, Wesley K., 1985. *The ultimate enemy : British intelligence and Nazi Germany, 1933-1939*, Cornell University Press.

Weinberg, Gerhard L., 1980. *The foreign policy of Hitler's Germany : starting World War II, 1937-1939*, University of Chicago Press.

Weisiger, Alex, 2013. *Logics of war : explanations for limited and unlimited conflicts*, Cornell University Press.

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

Yoder, Brandon K., 2019. Retrenchment as a screening mechanism: Power shifts, strategic withdrawal, and credible signals, *American Journal of Political Science*, **63**, 130–145.

Appendix

Table of Contents

A Formal Presentation of theory

A.1 Result 1

Result 1 is formally stated as follows.

Lemma A.1 *If*

$$\theta_n > \frac{\lambda(H-p) + k(1-\lambda) + w - \theta_i\rho_i}{\rho_n} \tag{1}$$

Then there is no PBE wherein $s^l(nr_1)$ appears on the path.

In the first period, the limited aims Challenger's expected value from second period competition is lower than peace. Thus, conjecture an equilibrium that includes $s^l(nr_1)$, $s^D(nc|nr_1, c|r_1)$. C_l can profitably deviate to r_1 if: $\theta_i\rho_i + \theta_n\rho_n - k + pH - w > \lambda(H-k)$. This rearranges to condition 1 as desired.

A.2 Equilibrium Analysis (results 2-4)

In what follows I solve for all the PBE in the model. They are all summarized in Table 4. Unfortunately, the clearest formal exposition does not neatly follow the presentation of results. First, I solve for the pure strategy separating equilibria. Second, I solved for the mixed strategy equilibria. I also rule out the possibility that other mixed strategy equilibria exist. Finally, I solve for pooling equilibria and the off-path beliefs necessary to sustain it under D2.

The analysis takes as given draws θ_i, θ_n. This is without loss of generality because these draws are publicly observed before the first action. The equilibrium we solve for cover the complete parameter space. Thus, an equilibrium takes all the below propositions and maps them onto specific draws of θ_i, θ_n.

Finally, all the analysis exploits the results of Lemma 2.1. we assume $\mathcal{A}_1, \mathcal{A}_2$, and take as proven $C's$ second period strategy. Further, we take as proven $D's$ competition choice given α_1 relative to α^*.

Pure strategy, separating equilibria We start with pure strategy, separating equilibria.

Proposition A.2 *If*

$$\theta_i\rho_i + \theta_n\rho_n > k(1+\lambda) - H\lambda(1-p) - w \tag{2}$$

$$\theta_i\pi_i + \theta_n\pi_n < 2k - H(1-p) - w \tag{3}$$

there is a fully informative violent trust equilibrium: $s^g(nr_1, r_2), s^l(r_1, r_2|\lambda = 1, nr_2|\lambda = 0), s^D(c|nr_1, nc|r_1)$. $\alpha_1|r_1 = 1, \alpha_1|nr_1 = 0$.

If instead,

$$\theta_i\rho_i + \theta_n\rho_n < H\lambda(1-p) + k(1-\lambda) + w \tag{4}$$

$$\theta_i\pi_i + \theta_n\pi_n > H(1-p) + w \tag{5}$$

there is a fully informative violent mistrust equilibrium: $s^g(r_1, r_2), s^l(nr_1, r_2|\lambda = 1, nr_2|\lambda = 0), s^D(nc|nr_1, c|r_1)$. $\alpha_1|r_1 = 0, \alpha_1|nr_1 = 1$.

By $\mathcal{A}_2$, D strictly prefers competition if $\alpha_1 = 0$ and peace if $\alpha_1 = 1$. The remainder of the proof are simply utility comparisons for C. Using the expected utilities reported in Table 2, the equilibrium conditions for violent trust contrasts utilities from row 3 and 6. Condition 2 states C^l prefers 3 over 6. Condition 3 states C^g prefers 6 over 3. D can observe all actions in equilibrium. Thus, there are no off-path beliefs. The equilibrium conditions for violent mis-trust contrasts utilities from row 2 and 5. Condition 4 states C^l

prefers 2 over 5. Condition 5 states C^g prefers 5 over 2. D can observe all actions in equilibrium. Thus, there are no off-path beliefs.

Notably, there cannot be any other complete separating equilibrium. If there were, either D would need to compete against C_l and avoid competition against C_g. This is not supported under $\mathcal{A}_2$.

Tangent: Can violent trust survive given C_g values issues more than C_l Several scholars have wondered if violent trust only survives when the limited aims Challenger values the issue more than the greedy Challenger. To be clear, state-types depend on their sensitivities to different sources of value. Recall I explicitly assumed that C^g always had more overall sensitivity than C^l, and C^g strictly cared about instrumental value more than normative value. This is a different question. It asks whether violent trust can survive when $\theta_i\pi_i + \theta_n\pi_n > \theta_i\rho_i + \theta_n\rho_n$.

Remark Suppose that the sensitivities and salience mean that C^g values the first issue more than C^l does. Then violent trust is still possible so long as $k > H(1-p)$.

This follows by checking if the RHS of 3 is greater than the RHS of 2. The condition implies that the first period issue is high stakes relative to the expected value of future issues for both states (that is, H is sufficiently low). To be clear, I deliberately omit a long discussion of this condition in the manuscript, or a broader analysis of the sensitivity parameters. The reason is we do not observe sensitivities in real life, or how states accrue utility from the interaction between salience and similarity (Kertzer 2016).

Mixed strategy, semi-separating equilibria. First, we solve for mixed strategy equilibria when $\alpha_0 > \alpha^*$. Notice we can only support equilibria where C^l mixes. The reason is that the mixing condition has to cause $\alpha_1 = \alpha^*$. There are two cases to consider: $s^g(r_1), s^g(nr_1)$. Starting with the first case. Let

$$\delta_1^* = \frac{\theta_i\rho_i + \theta_n\rho_n - k}{H\lambda(1-p) + w - k\lambda}$$

$$\gamma^* = \frac{(1-\alpha_0)(1-p-w)}{\alpha_0(p+w-\lambda)}$$

Proposition A.3 *If* $\gamma^* \in (0,1)$ *and*

$$0 < \delta_1^* < \frac{\theta_i\pi_i + \theta_n\pi_n - k}{H(1-p) + w - k}$$

Then the following mixed strategies are PBE. $s^D(nc|nr_1, pr(a = c|r_1) = \delta^*, pr(a = nc|r_1) = 1 - \delta^*), s^g(r1, r_2), s^l(pr(a_1 = r_1) = \gamma^*, pr(a_1 = nr_1) = 1 - \gamma^*, r_2|\lambda)$.

C^l is indifferent between revision and not if: $\theta_i\rho_i + \theta_n\rho_n - k + \delta(\lambda Hp - w) + (1-\delta)\lambda(H-k) > \lambda(H-k)$. This solves for δ_1^*. The equilibrium condition on δ ensures that C^g prefers $r_1|\delta_1^*$ to no revision.

D is indifferent between competition and not if, $\alpha_1|r_1 = \alpha^*$. This happens if after observing revision,

$$1 - p - w > \frac{\alpha_0\gamma}{\alpha_0\gamma + 1 - \alpha_0}(1-\lambda)$$

This solves for γ_1^* as desired. Trivially, $\alpha_1 = 0|nr$ as desired. There are no off-path beliefs to consider.

Turning to the second case, let

$$\delta_2^* = \frac{k - \theta_i\rho_i - \theta_n\rho_n}{H\lambda(1-p) + w - \lambda k}$$

Proposition A.4 *If* $\gamma^\dagger \in (0,1)$ *and*

$$0 < \delta_2^* < \frac{k - \theta_i\pi_i - \theta_n\pi_n}{H(1-p) + w - k}$$

Then the following mixed strategies are PBE. $s^D(nc|r_1, pr(a = c|nr_1) = \delta_2^*, pr(a = nc|nr_1) = 1 - \delta_2^*), s^g(nr1, r_2), s^l(pr(a_1 = nr_1) = \gamma^*, pr(a_1 = r_1) = 1 - \gamma^*, r_2|\lambda)$.

C^l is indifferent between revision and not if: $\theta_i\rho_i + \theta_n\rho_n - k + \lambda(H-k) < \delta(\lambda Hp - w) + (1-\delta)\lambda(H-k)$. This solves for δ_2^*. The equilibrium condition on δ ensures that C^g prefers $nr_1|\delta_2^*$ to no revision.

D is indifferent between competition and not if, $\alpha_1|nr_1 = \alpha^*$. This solves for γ^* as desired. Trivially, $\alpha_1 = 0|r_1$ as desired. There are no off-path beliefs to consider.

We now turn to mixed strategy equilibria when $\alpha_0 < \alpha^*$. Notice we can only support equilibria where C^g mixes. There are two cases to consider: $s^l(r_1), s^l(nr_1)$. Starting with the first case. Let

$$\delta_1^\dagger = \frac{\theta_i\pi_i + \theta_n\pi_n - 2k + H(1-p) + w}{H(1-p) + w - k}$$

$$\gamma^\dagger = \frac{\alpha_0(p + w - \lambda)}{(1-\alpha_0)(1-p-w)} = 1/\gamma^*$$

Proposition A.5 *If* $\gamma^* \in (0,1)$ *and*

$$0 < \delta_1^\dagger < \frac{\theta_i\rho_i + \theta_n\rho_n + w - k(1+\lambda) + H\lambda(1-p)}{\lambda H(1-p) + w - k\lambda}$$

Then the following mixed strategies are PBE. $s^D(c|nr_1, pr(a = c|r_1) = \delta_1^\dagger, pr(a = nc|r_1) = 1 - \delta^\dagger), s^l(r1, r_2|\lambda), s^g(pr(a_1 = r_1) = \gamma^\dagger, pr(a_1 = nr_1) = 1 - \gamma^\dagger, r_2)$.

C^g is indifferent between revision and not if: $\theta_i\pi_i + \theta_n\pi_n - k + \delta(pH - w) + (1-\delta)(H-k) = Hp - w$. This solves for $\delta_1^\dagger$. The equilibrium condition on $\delta_1^\dagger$ ensures that C^l prefers $r_1|\delta_1^\dagger$ to no revision.

Clearly, $\alpha_1|nr_1 = 0$, as desired. $\alpha_1|r_1 = \frac{\alpha_0}{\alpha_0 + (1-\alpha_0)\gamma}$. $\gamma^\dagger$ leaves D indifferent.

Let

$$\delta_2^\dagger = \frac{H(1-p) + w - \theta_i\pi_i - \theta_n\pi_n}{H(1-p) + w - k}$$

Proposition A.6 *If* $\gamma^\dagger \in (0,1)$ *and*

$$0 < \delta_2^\dagger < \frac{\lambda H(1-p) + w + k(1-\lambda) - \theta_i\rho_i - \theta_n\rho_n}{\lambda H(1-p) + w - \lambda k}$$

Then the following mixed strategies are PBE. $s^D(c|r_1, pr(a = c|nr_1) = \delta_1^\dagger, pr(a = nc|nr_1) = 1 - \delta^\dagger), s^l(nr1, r_2|\lambda), s^g(pr(a_1 = r_1) = \gamma^\dagger, pr(a_1 = nr_1) = 1 - \gamma^\dagger, r_2)$.

C^g is indifferent between revision and not if: $\theta_i\pi_i + \theta_n\pi_n - k + \delta(pH - w) + (1-\delta)(H-k) = Hp - w$. This solves for $\delta_1^\dagger$. The equilibrium condition on $\delta_1^\dagger$ ensures that C^l prefers $r_1|\delta_1^\dagger$ to no revision.

Clearly, $\alpha_1|nr_1 = 0$, as desired. $\alpha_1|r_1 = \frac{\alpha_0}{\alpha_0 + (1-\alpha_0)\gamma}$. $\gamma^\dagger$ leaves D indifferent.

Pooling equilibria Here we characterize four pooling equilibria labeled (1)-(4). All of these equilibria require off-path beliefs. We state the conditions under which they survive the intuitive criterion (IC). But later, we characterize the conditions we can support them under any set of prior beliefs.

Proposition A.7 *Assume $\alpha_0 > \alpha^*$, then if (1) condition 3 is violated and $\rho_i\theta_i + \rho_n\theta_n < k$ holds, then the following strategies form a PBE that survive the IC: $s^g(nr_1, r_2), s^l(nr_1, r_2|\lambda = 1, nr_2|\lambda = 0), s^D(nc|nr_1)$, and some off path belief that satisfies $\alpha_1 \geq \alpha_0|nr_1$. These PBE represent a status quo result.*

If, instead (2) $\rho_i\theta_i + \rho_n\theta_n > k$ and $\pi_i\theta_i + \pi_n\theta_n > k$ holds, then the following strategies form a PBE that survive the IC: $s^g(r_1, r_2), s^l(r_1, r_2|\lambda = 1, nr_2|\lambda = 0), s^D(nc|r_1)$, and some off-path belief that satisfies $\alpha_1 \geq \alpha_0|nr_1$. These PBE represent violent indifference.

Now assume $\alpha_0 < \alpha^*$, if (3) $\rho_i\theta_i + \rho_n\theta_n < k$ and $\pi_i\theta_i + \pi_n\theta_n < k$ holds and conditions 2 and 3 are jointly violated, then the following strategies form a PBE that survive the IC: $s^g(nr_1, r_2), s^l(nr_1, r_2|\lambda = 1, nr_2|\lambda = 0), s^D(c|nr_1)$, and some off-path belief that satisfies $\alpha_1 < \alpha_0|r_1$. These PBE represent a tragic status-quo.

If, instead (4) $\rho_i\theta_i + \rho_n\theta_n > k$ and $\pi_i\theta_i + \pi_n\theta_n > k$ holds, and 4 and 5 are jointly violated, then the following strategies then the following strategies form a PBE that survive the IC: $s^g(r_1, r_2), s^l(r_1, r_2|\lambda = 1, nr_2|\lambda = 0), s^D(c|r_1)$, and some off-path belief that satisfies $\alpha_1 < \alpha_0|nr_1$. These PBE represent the tragedy status-quo.

Since all Challenger-types pool, then D's posterior beliefs from on-path actions in every equilibrium are $\alpha_1 = \alpha_0$. Given Lemma 2.1, we can only support no competition, if $\alpha_0 > \alpha^*$, and competition otherwise, as desired.

To support equilibria (1) it must be that C^l does not want to deviate to revision. To start, consider the off path belief $\alpha_1|r_1 < \alpha^*$. In this case, all players prefer to remain on the path if condition 4 holds but 5 is violated. This sets the boundary for all class (1) equilibria. However, these equilibria degenerate under the intuitive criterion if $\rho_i\theta_i + \rho_n\theta_n > k$.

Consider the off-path belief $\alpha_1|r_1 \geq \alpha^*$. In this case, D plays no competition no matter what C does. C^l prefers no revision to revision if: $\rho_i\theta_i + \rho_n\theta_n - k + EU_2^l|s^D > EU_2^l|s^D$, where $EU_2^l|s^D$ represents C^l's second period expected utility holding D's strategy constant. C^g's comparison is defined similarly.

It follows that if $\rho_i\theta_i + \rho_n\theta_n > k$ holds, but $\pi_i\theta_i + \pi_n\theta_n > k$ does not. Then, C^l can profitably deviate to revision for the conjectured off-path belief, but C^g cannot. It follows that the belief $\alpha_1 = 1$ survives IC in these ranges. If true, then C^l profitably deviates to revision. If the reverse is true, then C^g can profitably deviate for the stated off-path belief but C^l cannot. It follows that $\alpha_1 = 0$ survives IC. If true, then neither player can profitably deviate to revision. If both hold, then neither holds an incentive to deviate. Thus, we can support off-path beliefs $\alpha_1 = \alpha^*$, as desired.

Turning to class (3) equilibria. Consider the off-path belief $\alpha_1|r_1 < \alpha^*$. In this case, the Challenger face competition no matter what he chooses. In this case, both Challengers strictly prefer to remain on the path (no revision) $\pi_i\theta_i + \pi_n\theta_n < k$, $\rho_i\theta_i + \rho_n\theta_n < k$. This sets the boundaries for class (3) equilibria. These equilibria degenerate under the intuitive criterion if conditions 2 and 3 hold (violent trust equilibria, 1a, in Figure 1). Consider the off-path beliefs $\alpha_1|r_1 \geq \alpha^*$. If true, then D plays competition if C plays revision. If conditions 2 and 3 hold, then C^l can profitably deviate to revision but C^g cannot given the conjectured off path belief. It follows that the belief $\alpha_1 = 1$ survives IC in these ranges. If true, then C^l profitably deviates to revision.

If conditions 2 holds, and 3 does not, then all players can profitable deviate to revision. It follows that the belief $\alpha_1 = \alpha_0$ survives IC. Within the equilibrium boundaries, no Challenger-type can profitably deviate. If conditions 2 and 3 are both violated, then only C^g can profitably deviate to revision. It follows that the belief $\alpha_1 = 0$ survives IC. Within the equilibrium boundaries, no Challenger-type can profitably deviate.

Turning to class (2) equilibria. Consider the off-path belief $\alpha_1|nr_1 < \alpha^*$. Then, D plays revision iff C plays no revision. In this case neither player can profitably deviate if condition 2 holds and 3 is violated.

This sets the boundaries for class 2 equilibria. However, these equilibria degenerate under IC if either $\rho_i\theta_i + \rho_n\theta_n < k$ or $\pi_i\theta_i + \pi_n\theta_n < k$ are violated. Consider the off-path belief $\alpha_1|nr_1 \geq \alpha^*$. In this case, D plays no competition no matter what C does. If $\rho_i\theta_i + \rho_n\theta_n > k$ and $\pi_i\theta_i + \pi_n\theta_n > k$, then neither Challenger profits form a deviation to nr_1. Thus, the belief $\alpha_1 = \alpha_0$ survives IC and no player can profitably deviate from class (2) equilibria.

Turning to class (4) equilibria. Consider the off-path belief $\alpha_1|nr_1 < \alpha^*$. Then D plays revision no matter what C does. Both players strictly prefer revision if $\rho_i\theta_i + \rho_n\theta_n > k$ and $\pi_i\theta_i + \pi_n\theta_n > k$ hold. This sets the boundary for class (4) equilibria. These equilibria degenerate under the intuitive criterion if conditions 4 and 5 hold (violent mistrust equilibria, 2a, in Figure 1). Consider the off-path beliefs $\alpha_1|nr_1 \geq \alpha^*$. If true, then D plays competition if C plays revision. If conditions 4 and 5 hold, then C^l can profitably deviate to no revision but C^g cannot given the conjectured off path belief. It follows that the belief $\alpha_1 = 1$ survives IC in these ranges. If true, then C^l profitably deviates to no revision.

A.3 Strategic selection into contests with observable endowments.

We make the following adjustments. At the first revision opportunities, C is presented with two issues to chose from. Thus, we re-define $a_1 \in \{nr, 1, 2\}$ where $a_1 = 1$ means C selected the first issue. C incurs k if C selects either revision opportunity.

For simplicity, we truncate the salience parameters. While this obscures some aspects of how D learns in cases, we will focus on cross-issue variation and strategic selection as a secondary mechanism. We assume that C_l values issue 1 l_1, and issue 2 l_2. We assume that C_g values issue 1 g_1, and issue 2 g_2. We focus on the interesting case where: $l_1 > l_2$, $l_1 > k$, $g_2 > g_1 > k$.

Substantively, this represents the following scenario. A series of strategic events have occurred that give C a plausible chance to revise the status quo over different issues. This could represent a setting where C borders two regions and can chose, which one to contest. It also could represent that at a specific moment, regional unrest/provocations towards C are occurring in two distinct theatres.[18] Based on the salience of these issues, the limited aims Challenger prefers issue 1 over 2 all else equal, and the greedy challenger prefers 2 over 1 all else equal. Finally, the limited aims type strictly prefers the value of revision over issues 1 (and only issue 1) over the cost. But C^g prefers broad revision.

A.3.1 Analysis

Since nothing changes in the second period, the results of Lemma 2.1 hold. Thus, we continue to assume $\mathcal{A}_1, \mathcal{A}_2$ and note the trust threshold α^*.

Because peace persists early in great power cases, we also focus on the case $\mathcal{A}_3 = \alpha_0 > \alpha^*$. This basically means, as in the Hitler and Stalin cases, that if British elites do not change their beliefs, that pace will persist.

We are most interested in the separating equilibrium.

Proposition A.8 *If $\mathcal{A}_1 - \mathcal{A}_3$ hold, and*

$$g_2 - g_1 > H(1-p) - k + w \tag{6}$$

then the following separating equilibrium survives the IC. In it, $s^D(nc|a_1 = 1; c|a_1 \in \{nr, 2\})$, $s^g(2, r_2)$, $s^l(1, r_2|\lambda = 1, nr_2|\lambda = 0)$.

[18] I have written a more general model that accounts for C selecting from J issues. The results are effectively the same because all that matters is that C_l selects her favourite issue, and C_g must select between C_g's favourite issue or pooling by selecting C_l's favourite issue. The model I present just focuses on those two choices. Owing to Appendix limits, and the limited additional insight tat follows from it, I do not present it.

In the baseline model, we solved for D's strategy and C's second period strategy. Thus, we focus on C's first period strategy. C^l on-path strategy generates the maximum possible utility. Thus, there is no deviation to consider. C_g has two potential deviations. First, C_g an deviate to $a_1 = 1$ and avoid competition. This deviation is profitable if $g_1 - k + H - k > g_2 - k + pH - w$. This re-arranges to the equilibrium condition. Because $g_2 > k$, this captures C^g's deviation to no revision. The off-path belief is that C is greedy if C avoids revision. Clearly, C_l must do worse from this than the on path strategy because C_1 accrues the maximum possible utility from on-path play. This completes the proof.

We now consider a pooling equilibrium.

Proposition A.9 *If $\mathcal{A}_1 - \mathcal{A}_3$ hold, and condition A.8 is violated then the following pooling equilibrium survives a D1 refinement. In it, $s^D(nc|a_1 = 1; cotherwise, s^g(1, r_2), s^l(1, r_2|\lambda = 1, nr_2|\lambda = 0)$. Off path, if $a_1 \neq 1$, D infers $\alpha_1 = 0$*

The difference is that violating A.8 ensures C_g strictly prefers to contests issue 1 over issue 2, given that it generates peace. Clearly, C_l cannot profitably deviate from her maximum value.

What would it take for the classic result to bare out. That is, trust is maximized when C avoids revision. We can achieve it if we assume that $l_1 < k$. In other words, there is no issue that C values more than the costs of a militaristic action.

A.4 Continuous types

I now adjust the baseline model to account for continuous variation in preferences. Define C's type as $\lambda \sim f[0, 1]$. Where $f()$ is non-negative, continuous and differentiable on $[0, 1]$. Assume that C's first period value for revision $U^c(r_1) = \theta_i\lambda + \theta_n(1 - \lambda)$. Further assume that Nature draws the second issue as high valued with probability $pr(\pi_2 = H) = x\lambda$. Here $x \in (0, 1)$ is a resolution parameter that allows for variation in the scope of greedy relative to limited preferences. Putting these features together means that as types are more sensitive to instrumental value, they are more likely to hold a high value for the second period issue. I assume C's type is private, and D only knows the distribution $f()$.

A.4.1 Analysis

The purpose of the analysis is to show that continuity in types does not ruin my main predictions. Thus, I solve the model for pure strategies semi-separating PBE that establish the conditions for violent trust and mistrust.[19]

The main difficulty with continuous types is defining increases in trust. We define it in terms of expected values of λ. Define, the prior level of trust as: $\alpha_0 = E(\lambda|f)$. Because of a notational change, higher values indicate less trust. Define $\alpha_1 = E(\lambda|s^g, s^l, a_1, f())$ as trust conditional on violence.

We continue to assume $\mathcal{A}_1$, and now replace $\mathcal{A}_2 = \frac{p+w}{x} < 1$. We also re-define $\alpha_1^* = \frac{p+w}{x}$. This gives us the following result

Lemma A.10 ***Bounding parameters to guarantee a trust problem.*** *If $\mathcal{A}_1, \mathcal{A}_2$ hold then the following strategies are on path in every Pure Bayesian equilibrium. If we arrive at the second revision opportunity, $s^g(r_2), s^l(r_2|\lambda = 1, nr_2|\lambda = 0)$. D's equilibrium decision to compete hinges on $s^A(c|\alpha_1 < \alpha^*, nc|\alpha_1 > \alpha^*)$.*

See the proof of the original Lemma.

We start by focusing on the separating equilibrium. We say an equilibrium is contingent if D's competition choice is conditional on C's strategy. Define a cut-point on C's type:

[19] Pooling equilibria also emerge when $\theta_i - \theta_n$ are sufficiently close, and k is either high or low.

$$z_1 = \frac{\theta_n + w - k}{\theta_n - \theta_l - x(H(1-p) - k)}$$

Proposition A.11 ***Contingent, violent trust.*** *If* $z_1 \in (0,1)$ *and*

$$\int_0^{z1} f(\lambda)d\lambda < \alpha^* < \int_{z1}^1 f(\lambda)d\lambda$$

Then the following is a pure strategy PBE. $s^D(nc|r_1, c|nr_1)$, $s^C(r_1|\lambda \leq z_1, nr_1|\lambda > z_1)$.

Given $s^D()$, C prefers revision to not if $\theta_i\lambda + \theta_n(1-\lambda) - k + x\lambda(H-k) > x\lambda Hp - w$. subbing in $z_1 = \lambda$ is the point of indifference. It follows, that C prefers revision in the first period if $\lambda \leq z_1$ and not otherwise.

As stated in A.10, D prefers competition if $\alpha_1 < \frac{p+w}{x}$ and not otherwise. $\int_z^1 f(\lambda)d\lambda = \alpha_1|s^C, f, nr_1$, $\int_0^z f(\lambda)d\lambda = \alpha_1|s^C, f, r_1$. There are no off path beliefs to consider.

Remark In equilibrium, first period revision raises trust and avoiding revision decreases trust.

Because $f()$ represents a pdf, $\alpha_1|r_1 > \alpha_1|nr_1$ given C's strategy. All that's left to do is show we can find $f()$ that supports this strategy pair. Note that the baseline model is a special case of the general $f()$. Clearly, there is a polynomial approximation of the discrete type space that satisfies this condition.

We now solve for the standard violent mistrust equilibrium. Define a cut-point on C's type:

$$z_2 = \frac{w + k - \theta_n}{\theta_l - \theta_n - x(H(1-p) - k)}$$

Then

Proposition A.12 ***Contingent, violent mistrust.*** *If* $z_2 \in (0,1)$ *and*

$$\int_0^{z2} f(\lambda)d\lambda < \alpha^* < \int_{z2}^1 f(\lambda)d\lambda$$

Then the following is a pure strategy, violent trust PBE. $s^D(c|nr_1, nc|r_1)$, $s^C(r_1|\lambda > z_1, nr_1|\lambda \leq z_1)$.

Given $s^D()$, C prefers revision to not if $\theta_i\lambda + \theta_n(1-\lambda) - k + x\lambda Hp - w > x\lambda(H-k)$. Subbing in $z_2 = \lambda$ is the point of indifference. It follows, that C prefers revision in the first period if $\lambda > z_2$ and not otherwise.

As stated in A.10, D prefers competition if $\alpha_1 < \frac{p+w}{x}$ and not otherwise. $\int_z^1 f d\lambda = \alpha_1|s^C, f, nr_1$, $\int_0^z f d\lambda = \alpha_1|s^C, f, r_1$. There are no off path beliefs to consider.

Remark In equilibrium, first period revision reduces trust and avoiding revision increases trust.

Because $f()$ represents a pdf, $\alpha_1|r_1 < \alpha_1|nr_1$ given C's strategy.

A.5 Situation-specific uncertainty

We adjust the baseline model as follows. In the first period, Nature selects a crisis-specific error $x \sim z[0, \bar{x}]$ and reveals it privately to C. If C selects revision in the first period, C accrue an additional x. Otherwise the model is the same.

I do not alter the structure of D's choice or second period payoffs. Thus, $\mathcal{A}_1, \mathcal{A}_2, \alpha^*$ all continue to apply.

The purpose of the extension is to show that we can still support our most interesting results when C has private information over the situation-specific variables. For simplicity, we focus on the parameter ranges that satisfy the fully informative, pure strategy, separating violent trust equilibrium in the baseline model. This is the first equilibrium described in proposition A.2. To refresh your memory, D plays $s^D(c|nr_1, nc|r_1)$. C's first period on-path strategy is $s^l(r_1, .), s^g(nr_1)$.

We can re-state C_l's ICC (conditions 2) including x as:

$$\theta_i\rho_i + \theta_n\rho_n + x > k(1+\lambda) - H\lambda(1-p) - w$$

true for any $x \geq 0$.

We can re-state C_g's ICC (conditions 3) including x as

$$\theta_i\pi_i + \theta_n\pi_n + x < 2k - H(1-p) - w$$

true iff

$$x < 2k - H(1-p) - w - \theta_i\pi_i - \theta_n\pi_n := x^*$$

If this condition is violated, then C^g can profitably deviate from nr_1 to r_1.

Thus, we can support the same basic equilibrium strategy with the following amendments. First, we redefine C^g's equilibrium strategy so $s^g(r_1|x \geq x^*, nr_1|x < x^*, r_2)$. Second, we impose a second condition on the equilibrium.

$$\frac{\alpha_0(p+w-\lambda)}{(1-p-w)(1-\alpha_0)} > \int_{x^*}^{\bar{x}} z(x)dx$$

We derive this from D's posterior belief: $\alpha_1 = \frac{\alpha_0}{\alpha_0 + (1-\alpha_0)\int_{x^*}^{\bar{x}} z(x)dx}$. And solving for $\alpha_1 > \alpha^*$.

A.6 Ambiguous Core interests

So far, Defenders have known whether an issue is highly salient for normative or instrumental reasons. In real life, there will be some ambiguity over this question. Before I study the model with ambiguity, I want to provide more substantive context for this assumption. My historical review of cases shows that Defenders frequently compile detailed lists of a Challenger's declared core interests (the territories not listed are implicitly peripheral interests). For example, in 1942, Britain's Strategic Intelligence Services provided a precise list of the issues and territories Stalin would care about if he was truly highly sensitive to the fear of foreign invasion. They compiled this list based on a review of Stalin's diplomatic statements and Russian history. This list suggested that dismembering Germany, and Soviet control over Poland and Eastern Europe fell inside Stalin's declared core interests. However, control over Turkey and Iran did not (i.e. where clearly peripheral). Similarly, Britain compiled a list of Prussian[20] (1850s) and German (1930) core interest claims. Britain also used the Monroe Doctrine to compile a list of American core interests. Similarly, at the end of the Cold War, president Bush ordered a National Security Review of all regional threats in the post-Cold War wold. The study team analyzed the core interest claims of India, Iraq, Iran, China, Russia, Turkey and many other middle powers. It made high-confidence assessments on the core interest claims of each state. But it could not assess with high confidence if the claims of Iraq, Russia and China were genuine.[21] Consistent

[20] There was some ambiguity over whether Austria was part of Prussia's core interests. Otherwise, the British knew which territories Prussia sought as part of unification.

[21] Author's interview with Former Deputy Secretary Amb. Robert Kimmit; who led the assessment.

with my theory, they did not know if these states held limited aims, but could assess with high confidence what each state would want if held limited aims.

Consistent with these cases, and with theories of international relations and intelligence, Defenders acquire this information through two sources. First, Defenders employ intelligence services to study the Challenger's history and culture (Lowenthal 2019). Experts know know where the Challenger's ethnic diaspora live, and the Challenger's historical borders and important religious sites. Second, Challengers almost always declare their core interests through Defense White Papers, public speeches and diplomacy long before they embark on periods of revision. Given what we know about incentives to coordinate (Trager 2010), it is reasonable to focus on these claims as the set of issues that the Challenger would value high if the Challenger held limited aims.

For the issues that Defenders can confidently list as core or peripheral interests my baseline model well fits. However, as these examples show, some specific issues remain ambiguous either because core interests can slowly change over time (Barnett 1999), or greedy Challengers can exploit arcane historical episodes to generate a pre-text when an opportunity for revision arises. In cases like this, D may not know if an issue is genuinely a core interest or not.

A.6.1 Set up

I adjust the model as follows. First, assume θ_i is a fixed publicly known value, and $f_n() = pr(\theta_n = \theta_H) = x, pr(\theta_n = \theta_L) = 1 - x$, where $\theta_H > \theta_L$. Assume that $f_n()$ is known but the realization of θ_n is shown privately to C. Second, assume that once θ_n is drawn, Nature reveals its true value to D with probability $\psi > .5$ and reveals the wrong value to D with probability $1 - \psi < .5$.[22] C knows the distribution ψ but not what D observes. Finally, define $\hat{\theta} \in \{\theta_H, \theta_L\}$ as what D observes.

This represents the following setting. An issue presents a true level of core salience to C, C knows that level of core salience, but D is uncertain. This could arise because D is incompletely informed about C's history, or C's core values are not completely articulated at a level to determine if the issue-salience fits. For example, Prussian unification of Germanic territories could have included Austria, or not. It depends on whether you interpret Germanic linguistically, historically, ethnically, or based on certain cultural and religious definitions. $\theta_H - \theta_L$ represents the difference in the issue salience for the contested issue given what C's true core values are. D employs intelligence analysts who write private estimates to D's government. Thus, D has an estimate, but that estimate is not publicly shared. ψ represents the quality of D's intelligence services. Nothing changes in the second period, and so $\alpha^*, \mathcal{A}_1, \mathcal{A}_2$ continue to apply.

It turns out that this kind of uncertainty has nuanced implications that require a complete manuscript to explore. Here I perform two tasks. First, I use one example equilibrium to explore the boundaries of this kind of uncertainty for classic arguments about violent mistrust. Second, I use one example equilibrium to consider the implications of this ambiguity for my novel equilibria.

A.6.2 Analysis of the classic trust problem with Ambiguity

The classic intuition is that ambiguities over the source of value makes signaling hard under the assumption that costly military actions engender mistrust.

To focus on the classic argument we make the following parameter restrictions. First, we narrow our focus on parameter values where if $\theta_n = \theta_L$ was public, that we would arrive at the classic result (avoiding revision signals trust). That is the second equilibrium in proposition A.2.

To refresh your memory, on path strategies include, $s^g(r_1, .), s^l(nr_1), s^D(nc|nr_1, c|r_1)$.

We can re-write C_l's ICC (condition 4) as: $\theta_i\rho_i + \theta_L\rho_n < H\lambda(1-p) + k(1-\lambda) + w$, and C_g's (condition 5) as: $\theta_i\pi_i + \theta_L\pi_n > H(1-p) + w$. Where we substitute $\theta_n = \theta_L$.

[22] This restriction is without loss of generality. If the signal is wrong more than right, then D just flips the inference.

We can instantly observe two facts from these constraints. First, raising the value of θ_n cannot alter C_g's strategy. The reason is that C_g prefers revision even in the face of competition if $\theta_n = \theta_L$. Second, $\theta_H - \theta_L$ must be sufficiently large to violate C_l's ICC. Specifically, to guarantee the result is ruined it must be:

$$\theta_L < \frac{H\lambda(1-p) + k(1-\lambda) + w - \theta_i\rho_i}{\rho_n} < \theta_H \tag{7}$$

This substantiates our claim in the manuscript that the scope of the misunderstanding in salience must be large enough to shift C_l's preferences over strategies.

Second, we restrict our attention to the pooling equilibrium where C plays revision, and D responds with peace when $\theta_n = \theta_H$ is observed. That is equilibrium (2) reported in proposition A.7. To refresh your memory, this requires $\alpha_0 > \alpha^*$, and the ICCs $\rho_i\theta_i + \rho_n\theta_n > k$ and $\pi_i\theta_i + \pi_n\theta_n > k$. Subbing in $\theta_H = \theta_n$, the binding constraint is:

$$\frac{k - \rho_i\theta_i}{\rho_n} < \theta_H$$

We are now ready for the equilibrium analysis:

Proposition A.13 *If the above conditions are satisfied, and*

$$\psi > \frac{\alpha_0 x(p + w - \lambda) - 1 + \alpha_0}{\alpha_0 x(p + w - \lambda)} \tag{8}$$

then the following strategies are an SPE, $s^g(r_1, r_2)$, $s^l(nr_1|\theta_L, r_1|\theta_H, r_2|\lambda = 1, nr_2|\lambda = 0)$, $s^D(c|(r_1 \;\&\; \hat{\theta} = \theta_L); nc$ *otherwise*$)$. *There are no off-path beliefs.*

C^g's strategy is identical to the baseline. It imposes no additional constraints on the equilibrium. C_l's first-period strategy and D's beliefs are novel.

Starting with D's beliefs and strategies. On path, C_l never selects nr_1. Thus, $\alpha_1|nr_1 = 1$, as desired. There are two other cases to consider:

$$\alpha_1|r_1, \hat{\theta_H} = \frac{\alpha_0 x\psi}{\alpha_0 x\psi + 1 - \alpha_0}$$

D's on-path strategy is sustained if: $\frac{\alpha_0 x\psi}{\alpha_0 x\psi + 1 - \alpha_0} > \frac{1-p-w}{1-\lambda}$.

This re-arranges to $\frac{(1-p-w)(1-\alpha_0)}{\alpha_0(1-x)(p+w-\lambda)} > \psi$. But recall, $\alpha_0 > \alpha^*$. Subbing in the minimum bound, we arrive at: $\frac{1}{1-x} > \psi$. This is always true.

The second case is.

$$\alpha_1|r_1, \hat{\theta_L} = \frac{\alpha_0 x(1-\psi)}{\alpha_0 x(1-\psi) + 1 - \alpha_0}$$

D's on-path strategy is sustained if: $\frac{\alpha_0 x(1-\psi)}{\alpha_0 x(1-\psi) + 1 - \alpha_0} < \frac{1-p-w}{1-\lambda}$. This re-arranges to the equilibrium condition 8, as desired.

Turning to C_l's strategy. The θ_L case is trivial and imposes no additional equilibrium constraints. I focus, on the θ_H case. In this case, C_l prefers revision if:

$$\psi(\rho_i\theta_i + \rho_n\theta_H - k + \lambda(H-k)) + (1-\psi)(\rho_i\theta_i + \rho_n\theta_H - k + \lambda Hp - w) > \lambda(H-k)$$

$$\psi > \frac{w + k(1-\lambda) + \lambda(H(1-p) - \theta_i\rho_i - \theta_H\rho_n}{w + \lambda(H(1-p) - k)}$$

Notice the RHS must be negative by condition 7. Thus, it is always satisfied. There are no off-path beliefs.

Discussion of condition 8 Notice we can always satisfy condition 8 for x sufficiently low, and x appears in no other equilibrium constraints. Thus, we can satisfy this condition. When x is high enough, this condition is violated, and the equilibrium degenerates. However, it does not degenerate because of incentives for competition. Rather it degenerates because of D's incentives to select peace. Another equilibrium arises where D plays no competition unconditionally. This is clearly not what the intuitive arguments suggest.

Furthermore, we can support 8 at low levels of ψ. Subbing in $\psi = 1/2$, we arrive at:

$$2(1-\alpha_0)/\alpha_0 > x(p+w-\lambda)$$

This substantiates our claim that we can support it even at reasonably low levels of ψ.

A.6.3 Analysis of my novel equilibrium with ambiguity

We focus on the following parameter ranges. If D knew that $\theta = \theta_H$, then there would be a fully informative violent trust equilibrium. But if D knew $\theta = \theta_L$, there would be a pooling equilibrium on no revision, leading to competition. The conditions are $\alpha_0 < \alpha^*$,

$$\theta_H > \frac{k(1+\lambda) - H\lambda(1-p) - w - \theta_i\rho_i}{\rho_n} > \theta_L$$

$$\theta_H < \frac{2k - H(1-p) - w - \theta_i\pi_i}{\pi_n}$$

Proposition A.14 *If the abovementioned conditions are met, and*

$$\psi > \frac{(1-x)(p+w-\lambda)\alpha_0 - (1-p-w)(1-\alpha_0}{(1-x)(p+w-\lambda)\alpha_0} \tag{9}$$

then the following strategies are an SPE, $s^g(nr_1, r_2)$, $s^l(r_1|\theta_H, r_1|\theta_L, r_2|\lambda = 1, nr_2|\lambda = 0)$, $s^D(c|nr_1, nc|r_1)$. *There are no off-path beliefs.*

Starting with D's beliefs given C's actions.

Remark Taking a costly military action continued to generate trust: $\alpha_1|r_1 = 1 > \alpha_0$.

Since only the limited aims Challenger ever selects revision, ambiguity does not alter the result.

Avoiding a costly military action continues to engender mistrust, but the reduction in trust is diminished by ambiguity.

$$\alpha_1|nr_1, \hat{\theta} = \theta_L = \frac{\alpha_0(1-x)(1-\psi)}{\alpha_0(1-x)(1-\psi) + 1 - \alpha} > 0$$

For the equilibrium to hold together, it must be that $\alpha_1 < \alpha^*$. This re-arranges to the equilibrium condition as desired.

$$\alpha_1|nr_1, \hat{\theta} = \theta_H = \frac{\alpha_0 x\psi}{\alpha_0 x\psi + 1 - \alpha} > 0$$

For the equilibrium to hold together, it must be that $\alpha_1 < \alpha^*$. This is strictly true.

C's strategies are trivial given the initial conditions.

Finally, we remark on a peculiar result. Define β represent B's expectation that $\theta = \theta_H$.

Remark Costly military action weakly raise D's expectation that an issue is salient to core values. In particular, $\beta|\hat{\theta} = \theta_L, r_1 = 1 > \beta|\hat{\theta} = \theta_L$.

Ordinarily, in trust models we think about information as a one-directional problem where we infer C's intrinsic motivations from C's actions and the state of the world. Thus, ambiguity over the state of the world, makes D's problem harder. Here we see that information flows in two directions. It is possible that we can infer information that resolves ambiguity, based on knowledge of what C_l will do in equilibrium.