

# Violent Trust

June 1, 2022

## **Abstract**

How do costly, military choices affect great power competition? I integrate historical and normative preferences into a strategic theory of trust and great power competition. Against the conventional wisdom, I find that costly military actions, (e.g. invasions) signal limited and not expansive strategic intentions. Violence engenders trust because limited aims Challengers are sensitive to historical and cultural value and therefore care intensely about a handful of issues. Fighting for historically salient territories reveals limited aims because limited aims Challengers care so much they fight even when the cost of war is high. I use survey experiments to confirm that violent revision over a core interest engenders trust and avoiding revision engenders mistrust. The results move normative, domestic and rationalist approaches towards a unified framework of great power competition, clarify unappreciated differences between formal and informal rationalist arguments of great power trust, and illuminate Anglo-German (1930s) and Sino-American interactions.

Word Count: 10,890

## Table of Contents

---

<b>1</b>	<b>The strategic logic of trust</b>	<b>4</b>
1.1	Grounding limited aims in historical and cultural value . . . . .	7
1.2	Strategic Implications . . . . .	11
<b>2</b>	<b>Experimental evidence</b>	<b>18</b>
2.1	Vignette Structure . . . . .	19
2.2	Measuring Trust . . . . .	21
2.3	Pre-registered Hypotheses . . . . .	22
2.4	Results . . . . .	23
2.5	Second (abstract) study . . . . .	27
<b>3</b>	<b>Anecdotes and puzzling quantitative results</b>	<b>29</b>
<b>4</b>	<b>Conclusion</b>	<b>31</b>
	<b>Appendix</b>	<b>38</b>

---

In 1995, China sought to revise the status quo over Taiwan. China initiated an invasion plan, moved 150,000 forces to its coastline, and executed live-fire drills. China backed down only after the United States sailed a carrier through the Taiwan Strait. But even then China did not stop. In 1996, China invested in offensive battleships and used live-fire exercises to influence Taiwan's election. American foreign policy elites acknowledged that they had underestimated China's interest in violent territorial revision over Taiwan.<sup>1</sup> Surprisingly, several analysts also raised their confidence that China's long-term intentions were "limited to peaceful reunification" (Qimao 1996). The 1996 National Security Strategy shifted US policy towards Sino-American cooperation. For the first time it stated, "We have adopted a policy of comprehensive engagement designed to integrate China into the international community as a responsible member and to foster bilateral cooperation in areas of common interest."<sup>2</sup>

This episode should puzzle scholars of trust (Kydd 2005), great power competition (Waltz 1979), and the security dilemma (Jervis 1978). If anything, scholars think that violent territorial demands (Trager 2016), or either rapid or offensive arming signal aggressive motives (Glaser 2010; Coe and Vaynman 2019). The mistrust that follows causes great power competition to intensify (Braumoeller 2008), leading to catastrophic competition (Weisiger 2013). In this case we see the opposite. After China threatened war and invested in offensive weapons, Clinton deepened Sino-American commercial ties because he believed long-term cooperation was plausible.<sup>3</sup>

Can violent military actions engender trust? Answering the call of Fearon and Wendt (2002), and consistent with the domestic micro-foundations of state-preferences (Kertzer 2017), I embed insights about the historical and cultural determinants of state-preferences into a rationalist model of trust. I assume that limited aims Challengers are sensitive to historical and cultural value of territory (Abramson and Carter 2016; Goemans and Schultz 2017). They hold limited aims because they care intensely about a handful of core interests that hold historical and cultural significance to them. In contrast, greedy Challengers are often less sensitive to their unique historical and cultural context. They often pursue opportunistic revision when instrumental benefits outweigh instrumental cost (Coe and Markowitz 2021).

To be clear, many rationalist scholars informally describe limited aims states as motivated by non-

---

<sup>1</sup>A declassified assessment on China's 1996 military exercises from ONI, affirmed by the CIA, reads 'China is likely to conduct similar, politically motivated exercises in the vicinity of Taiwan in the near future.'

<sup>2</sup>In contrast, the 1995 NSS reads, "We are developing a broader engagement with the People's Republic of China."

<sup>3</sup>Some argue that the Taiwan Crisis caused analysts to estimate China's future preferences could be changed through enmeshment. The changing-preferences interpretation, which is also inexplicable for rationalist trust theory, compliments my theory because it means that perceptions of long-term cooperation followed violent actions. My elite interviews also confirm that hopes of shifting preferences was not the dominant cause of optimism.

material preferences (Glaser and Kaufmann 1998). However, it is not how formal presentations, from which informal scholars derive core insights, model them. Game theorists assume that limited aims Challengers value all issues less than greedy Challengers in expectation (eg Kydd 2005). I manipulate the type-space so limited aims Challengers value a few core interests as much or more than greedy Challengers. This subtle difference can reverse predictions about the relationship between military actions, reputation, and patterns of great power competition derived from standard signaling models (see Ramsay 2017).

I argue that Challengers can rapidly militarize, violate arms control agreements and even invade territories to capture issues of historical and cultural value without engendering mistrust. The intuitive logic is that Defenders expect limited aims Challengers to build offensive weapons and take territory that serve their core interests. Using a formal analysis I push my argument further. I show that violent territorial revision engenders trust if the Challenger is faced with the opportunity to pursue revision over an issue that has large historical and cultural value relative to its instrumental value. The logic is as follows. Revision is always costly. In situations where the instrumental costs of revision are high, opportunistic Challengers who are sensitive to instrumental value prefer to avoid revision. However, Challengers who are sensitive to historical and cultural value will revise the status quo at high instrumental cost if the issue in dispute has high historical or cultural value.<sup>4</sup> Defenders don't know all the situation-specific costs. But they can often identify specific territories (like Taiwan) that hold large historical value. When Challengers fight for territory of high historical value, they signal that they are unusually sensitive to this value. In contrast, Challengers that avoid revision over territories of historical and cultural importance communicate that they are more sensitive to instrumental costs and benefits. In doing so, they (probabilistically) reveal themselves as opportunistic types that will fight whenever the instrumental costs outweigh the benefits.

Many territories are valuable for both instrumental and normative reasons. Others argue that this ambiguity makes learning difficult (Mearsheimer 2001). I show theoretically that when Challengers fight for ambiguous issues it has no affect trust, and therefore no affect on the Defender's choice between competition and peace.

I test my most surprising predictions using a survey experiment that measures beliefs (not preferences) (like Renshon, Dafoe, and Huth 2018) in a trust scenario (like Kertzer and McGraw 2012). In it, I tell subjects about the emergence of a revisionist Japan in 2040. I ask subjects to assess if Japan holds expansive or limited aims. The experiment is unique because all treatments include identical actions: Japan uses

---

<sup>4</sup>I explicitly assume that each situation is unique, and Defenders only hold expectations about the Challenger's costs and benefits.

military force to take territory from Korea. However, I randomly assign whether Japan is faced with a crisis over an issue that unambiguously serves historical and cultural value (Amami Islands); ambiguously serves both historical and instrumental value (Senkaku Islands); or has only a small historical connection, but a moderate amount strategic value (Liancourt Island). In each treatment arm Japan explains that revision serves its historical objectives. However, factual information about the amount of historical and cultural context determines how subjects evaluate Japan's intentions. Subjects that observe Japan violently take the Amami Islands are more trusting of Japan than they were before the crisis, and more trusting than they would have been if they observed Japan takes the Liancourt Island. I use an abstract experiment to confirm my main results and show that fighting for a core interest engenders more trust than avoiding war over a core interest. I use anecdotes to illustrate how many puzzling but critical historical cases, e.g Hitler's revision in German-speaking Europe, plausibly fit my logic.

Putting it altogether, I find violence engenders trust and peace when contested issues hold clear historical importance. Violence has no affect on trust when contested issues hold both historical and instrumental value. In contrast, and consistent with conventional findings, violence engenders mistrust and great power competition if either: Challengers fight for peripheral interests; Defenders are unable to determine if the Challenger values a disputed territory for historical reasons with at least moderate confidence; or greedy Challengers value all issues more than limited aims Challengers in expectation.

Using this result, I show that we cannot draw simple linear predictions that rapid militarization, offensive arming, or territorial demands communicate aggressive intentions or high resolve (In addition to above, [Powell 1996](#); [Yoder 2019](#); [Debs and Monteiro 2014](#); [Carr 1945](#)). But I also show that context does not prevent states from drawing reliable inferences from past actions ([Mercer 1996](#); [Press 2007](#)) because a rival's historical context is somewhat observable. Defenders who are uncertain of a rival's true intentions, can still use historical knowledge to adjudicate how close an issue is to that rival's core interests.

By embedding historical context into a rationalist model of international competition, I forge a closer connection between those that identify the determinants of foreign policy orientations using either constructivism ([O'Neill 1999](#); [Finnemore 1996a](#)) or domestic politics ([Trager and Vavreck 2011](#); [Moravcsik 1997](#)), with rational accounts of great power competition. For constructivists, I provide a distinct normative logic that compliments ontological trust ([Mitzen 2006](#)). For domestic scholars of conflict I show how normative preference creates unexplored signaling opportunities because they impacts which territories states value ([Crescenzi 2007](#); [Trager 2010](#)), and not just the scope of preferences. This insight clarifies how normative

preferences, such as nationalism or historical control, interact with strategic choice to generate different (possibly opposite) patterns of war and peace in different historical contexts (Powers 2022; Jackson and Morelli 2011; Carter and Goemans 2011).

## 1 The strategic logic of trust

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), great power rivalries (Glaser 2010), arms control (Coe and Vaynman 2019; Vaynman 2021), economic cooperation (Svolik 2006), leadership targeting (McGillivray and Smith 2008), and general problems of deterrence (Gurantz and Hirsch 2017). The security dilemma (Jervis 1978) that drives the tension in realism, and the spiral that it can generate, are a two-sided trust problem (Haynes and K. Yoder 2020). The most famous applications of the repeated prison's dilemma with incomplete information are also trust problems (Axelrod 1980). Trust problems are the focus of American policy-makers. Many believe that Sino-American relations (Glaser 2015) and cyber-conflicts (Axelrod and Iliev 2014) are arguably problems of trust as much as resolve.

Each trust theory is built differently to account for its empirical domain. However, they all share two common features. First, they assume that states hold different strategic intentions (Glaser 1995). Some states have cooperative intentions (Jervis 1978). These types are labeled as security seekers (Mearsheimer 2001) or limited aims (Edelstein 2019). Other states have non-cooperative intentions. These types are labeled as greedy or expansive.

Second, these theories share a strategic tension. I present the simplest set of actions that accounts for that tension in Figure 1.<sup>5</sup> This simplistic model represents a one-sided trust problem between a status quo Defender and a revisionist Challenger.<sup>6</sup> The Challenger will soon have the opportunity to revise the status quo in her favor (step (4)). But before she does, the Defender has a (costly) opportunity to make it either more difficult or expensive for the Challenger to pursue revision (step (3)). This is described as a preventive action by deterrence scholars.

Like other commitment problems, if the Defender knows that the Challenger will revise the status quo in the future, he prefers to take a preventive action that makes it more difficult or expensive for the Challenger

---

<sup>5</sup>The sequence of moves is identical to the deterrence model reported by Gurantz and Hirsch (2017). I get different results because I model the Challenger's preferences and information about preferences differently.

<sup>6</sup>Each theoretical tradition complicates this problem. For example, realism assumes a two-sided trust problem.

Table 1: Simplest sequence of moves for trust theory

Step (1): Challenger's (C) Disposition	
	<p>Nature determines if the Challenger has greedy or limited aims (private for C). Usually modeled as variation in value for disputed issues (<math>v(\pi)</math>)</p>
Step (2): C's revision opportunity that could engender trust/mistrust.	
	<p>An opportunity for revision arises. Nature determines C's situation specific cost of revision <math>k_1</math> (private for C). Then, C decides to capitalize on that opportunity/not.</p> <p><b>Closest substantive interpretation:</b> C given opportunity to make territorial revision or orchestrate coup against a third-party state (i.e. not the Defender's sovereign territory, but a state that is of interest to the Defender).</p>
Step (3): Defender's (D) Opportunity for Preventive Competition	
	<p>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).</p> <p><b>Closest substantive interpretation:</b> Preventive war, Cold War style containment.</p>
Step (4): C's revision opportunity.	
	<p>Nature determines C's situation specific cost of revision <math>k_2</math> (private for C). Then, C decides to capitalize on that opportunity/not. See step (2) for closest substantive interpretation.</p>

*Note:* The sequence of moves is the minimum to generate the trust problem at the heart of research on trust, reassurance and the security dilemma. More complex models allow for two-sided uncertainty, C's arming or not arming (shifting  $p, k_2$ ), reciprocal arming, or allowing D to intervene at additional points. These complications do not ruin the results below.

to pursue revision tomorrow (Powell 2006). Unlike commitment problems, theories of trust assume that not all Challengers want to revise the status quo tomorrow. Rather, their interest in future revision depends on whether their preferences are greedy or limited (step (1)). The Defender is uncertain about the Challenger's preferences and therefore is uncertain if the Challenger wants to take the opportunity to revise the status quo tomorrow. If the Defender is sufficiently confident that the Challenger holds limited aims, he does not intervene in the hopes that the Challenger will cooperate in the future (avoids competition at step (3)). However, if he believes that the Challenger is likely greedy, he takes a preventive action to make future revision more difficult or expensive.

From the Challenger's perspective, these dynamics generate a reassurance problem (Haynes and K. Yoder 2020). All Challengers want to reassure the Defender that their aims are limited to avoid preventive competition. Defenders know that limited aims Challengers exist who will honestly promise that their aims are limited.

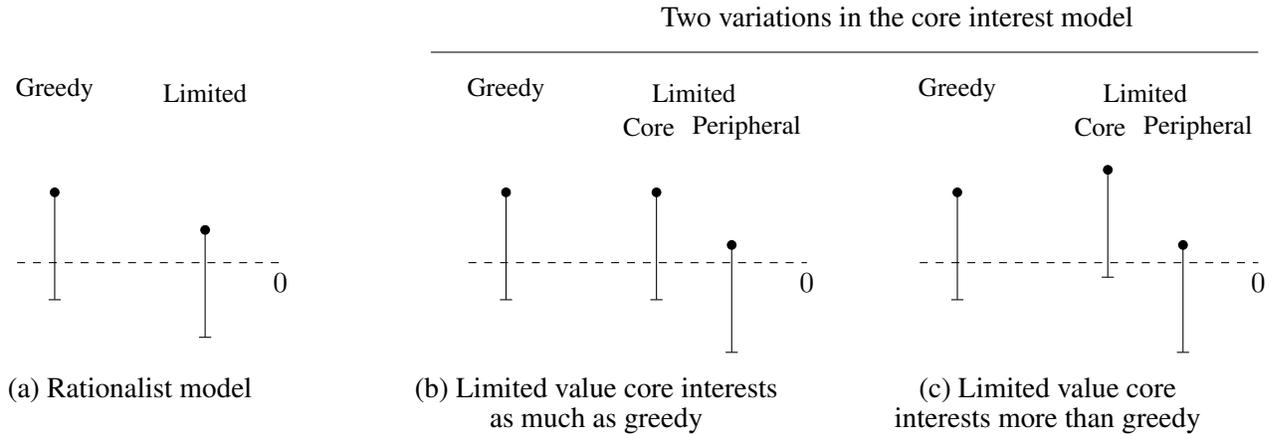
For a long time, scholars thought that problems of trust were insurmountable (Mearsheimer 2001). However, recent studies show that states can use costly signals to communicate limited aims (step (2)). Most rationalist studies focus on military or territorial signals (Kydd 2005; Coe and Vaynman 2019).<sup>7</sup> In these accounts, the Challenger is presented with an opportunity to militarize, take territory, or to decide whether to invest in offensive or defensive weapons. The limited aims Challenger credibly reveals her limited aims if she forgoes military investments, or territorial conquest, or offers concessions, or builds mainly defensive weapons, or commits to intrusive monitoring agreements. In contrast, greedy Challengers are tempted by the short term incentives to pursue revisionist policies. They reveal their type through militarizing or taking territory. This leads to the following conventional wisdom:

**Conventional Trust Prediction:** (a) When the Challenger is faced with an opportunity to engender trust through her military actions, she engenders mistrust if she: militarizes (especially with offensive weapons), takes territory or otherwise seeks to undermine the status quo. She engenders trust if she de-militarizes, gives concessions or otherwise avoids opportunities to undermine the status quo. (b) The more the Defender trusts the Challenger, the more likely it is that the Defender avoids preventive competition.

---

<sup>7</sup>Others focus on social explanations, diplomacy, or misperceptions. A handful of rational studies think about alliances. None of these studies disagree with the rationalist military signalling logic I characterize.

Figure 1: Operationalizing Motives



### 1.1 Grounding limited aims in historical and cultural value

As with all theories, rationalist theories of competition and trust simplify to focus on what they care about: the relationship between military actions, beliefs about the Challenger’s long-term intentions (i.e. trust) and great power competition. In what follows, I interrogate simplifying assumptions that rationalist make about how preferences vary.

To a large degree, I focus on formalized theories. Of course, many scholars use rigorous informal logic that do not explicitly make the assumptions below. However, informal scholars make these assumptions implicitly when they incorporate the signaling logic derived from formal theories. For example, Glaser (2010) explicitly cites Kydd (2007) to verify the logic of military signaling. He then argues that the same costly signaling mechanism applies in repeated interactions that determine grand strategic orientations. But he does not appreciate that this extension may not follow naturally if historical and cultural preferences determine these orientations for some states.<sup>8</sup>

Figure 1(a) visualizes how scholars typically operationalize variation in motives in formal models. The solid circles represent how much different kinds of Challengers benefit from seeking revision without factoring in the costs of revision. These solid circles are determined by the Challenger’s type (i.e are they are limited aims or greedy type).

The lines below the circles represent the Challenger’s potential net benefit from revision factoring in their value for revision (which is determined by their type) and the situational-specific cost of revision. The

<sup>8</sup>Because of these difficulties Glaser cannot easily explain trust during Anglo-American competition circa 1900, trust despite China’s claims for Taiwan, etc.

Challenger's costs of revision can vary from one situation to the next. This allow for potential variation in the Challenger's net value (i.e. benefits less costs) from revision versus no revision. Each point on the line represents a Challenger's cost-benefit calculation for revision given their motives (limited or greedy) and the situation-specific costs. If the point on the line is below 0, it means that the costs of revision outweigh the benefits in that situation for that type. If the point on the line is above 0 it means that the benefits outweigh the costs. As the plot shows, situation-specific costs create important variation in the Challenger's cost-benefit calculation. However, differences in each Challenger-type's absolute benefits ensure that Challengers with limited aims prefer to avoid revision for more situation-specific costs. In the mechanics of signaling theories, this assumption incentivizes limited aims Challengers to avoid revision in the first period because they are, in expectation, less sensitive to giving up on the first issue than greedy types. This, in turn, generates opportunities for limited aims Challengers to avoid revision to signal limited aims.

This simplifying assumption is appropriate if states are purely security-seekers or purely greedy. However, scholars who analyze the determinants of state preferences would describe limited aims in a different way (Finnemore 1996b; Katzenstein 1996).<sup>9</sup> I depict an operationalization that is closer to these findings in panels (b) and (c). In this presentation, greedy Challengers have a large value for revision over all issues. Limited aims Challengers care intensely about a handful of core interests. But their aims are limited because they have little value for the vast majority of issues (their peripheral interests) (Joseph 2020). It is possible that limited aims Challengers care about their core interests as much (b) or more (c) than greedy Challengers depending on the situation.

My strategic model abstracts away from debates about the specific microfoundations of greedy and limited aims in a particular case. For my strategic implications to follow, the most important thing is that the variation in state motives looks more like the core interest model of preferences presented in Figure 1(b) or (c) than the security-seeker model presented in 1(a). For ease of exposition, and consistent with the general thrust of literature on reputation (See Dafoe, Renshon, and Huth 2014, p. 382), I describe greedy types as those who are relatively sensitive to instrumental value and limited types as those who are relatively sensitive to historical and cultural value. But to appreciate the empirical implications of the model, which I validate using an experiment, it is useful to connect different parts of the literature on the determinants of state preferences to my type-space.

Research into the determinants of foreign policy preferences is ongoing. However, there are a few points

---

<sup>9</sup>It also does not match the discussions of limited aims in informal rationalist arguments (Frieden 1999).

of broad agreement that allow me to arrive at the abstractions presented in panels (b) and (c). First, each specific territory is unique (Mercer 1996). Therefore, different territories can be valuable to a Challenger for different reasons (Weisiger and Yarhi-Milo 2015). For example, some territories are associated with a state's history and culture (e.g. China's historical claims over Taiwan and Tibet). These territories provide normative value (Hopf 1994). Other territories are endowed with natural resources, productive populations or strategic positions that will allow for economic growth. These territories provide instrumental value.<sup>10</sup> Second, states through history have held different foreign policies because they are sensitive to different sources of value (Moravcsik 1997). This could be because of varying selectorates (Snyder 1993), leader-level preferences (Kertzer 2016), or international practice (Wendt 1992).<sup>11</sup>

Connecting these two insights together, whether a state holds limited or greedy foreign policy aims depends on how sensitive that state is to the different sources of value; and the prevalence of territories that serve those values. In many cases limited aims Challengers are highly sensitive to territories that they historically controlled (Carter and Goemans 2011), or that contain their ethnic and religious kin (Goemans and Schultz 2017). These values drive many states to hold limited aims because only a handful of issues serve a specific state's unique history in many cases. For example, around 1850, Prussia's goals of ethnic unification limited its goals to a few territories in Europe.

In contrast, many believe that greedy Challengers usually prioritize the instrumental value of revision (Glaser 2010). Thus, in many cases, greedy challengers see revision as a business proposition that weighs the economic or strategic value of revision against the strategic or economic costs.

My model accounts for three complicating factors that generate cross-case variation in the determinants of state preferences. First, scholars debate the sources of greedy preferences (see Coe and Markowitz 2021; Little and Zeitzoff 2017).<sup>12</sup> Certain norms can drive greedy aims such as status (Renshon 2016), certain variants of nationalism (see Powers 2022), or desires to spread ideology (Jackson and Morelli 2011). But if states covet regional domination, then they will prioritize territories that afford them the industrial base and strategic advantage to expand quickly. These territories may not contain ethnic kin, or cultural sites. My theory only assumes that limited aims Challengers value a peripheral interest less than a greedy Challenger values the same issue; and (2) core interest issues are relatively rare.<sup>13</sup>

---

<sup>10</sup>This description is not exhaustive. There are other sources of value, and many territories hold a mixture of different values.

<sup>11</sup>I am agnostic about the sources of value. My theory only requires that different states are sensitive to different values.

<sup>12</sup>Haynes and K. Yoder (2020) model compatible versus incompatible preferences. In contrast, I dis-aggregate territories from their values.

<sup>13</sup>I discuss these scope conditions, and the Russo-Ukraine War, in section 1.2.2.

Second, I do not assume that limited aims Challengers only care about historical and cultural value, and greedy Challengers only care about instrumental value. Rather, I only assume that limited aims Challengers can be more sensitive to values associated with their unique historical and cultural experiences and are less sensitive to other sources of value than greedy Challengers are.<sup>14</sup> As a result, when an issue in dispute involves questions of history and culture, the limited aims Challenger's value is activated to a larger degree in some cases.

Third, I accept that each territory is unique. Some core interest issues will hold huge amounts of historical value but little strategic value. For example, the Gaza Strip is a desert with no natural resources. The instrumental gain from this territory is very small. However, Israel and the Palestinian Authority care about it because of its historical importance. I say these are unambiguous core interests since limited types clearly value them more. This fits the modeling assumptions in Figure 1(c). Other core interests hold both large amounts of historical/cultural and instrumental value. For example, in 1924, Saudi Arabia claimed that the Kingdom of Hejaz was not properly caring for Mecca and Medina. Saudi Arabia invaded, claiming that protecting Muslim sites was in its core interest. But Hejaz also held extensive oil reserves and maritime access to the Strait of Hermouz. I say these are ambiguous core interests because both limited and greedy Challengers value them a lot. This fits the modeling assumptions in Figure 1(b). Many argue that ambiguity over the source of value can drive spirals of mistrust (Mearsheimer 2001). I explicitly test this claim in my strategic discussion.

My baseline model assumes that the Defenders know which issues fit the Challengers core and peripheral interests. To be clear, I still assume situation-specific uncertainty, and also explore ambiguity caused by variation in the relative historical and strategic value of core-interest issues. My assumption only means that the US knows that control over Taiwan and Tibet is consistent with China's history and culture and that control over any territory in Europe, Africa and South America are not. As the Director of National Intelligence explained to me during an interview, "Determining core interests has a long and honorable place in analysis. It is not that difficult [to do]. You ask any 50 China-analysts and they'd give you that same list of things that I came up with."<sup>15</sup> In Appendix A.7, I explain that real-life Defenders exploit cultural experts at intelligence agencies (Lowenthal 2019) and diplomatic interactions (Trager 2010) to acquire this information. But I also accept that some uncertainty likely remain. I extend the following model so that

---

<sup>14</sup>I explicitly test this assumption later using survey.

<sup>15</sup>Author's interview with Denis Blair.

the Defender is uncertain if an issue fits the Challenger's core or peripheral interests. The extension reveals several nuances but does not change more core findings so long as Defender's can list core interest with moderate confidence.<sup>16</sup>

## 1.2 Strategic Implications

Putting it altogether, I argue that what separates core and peripheral interests is that core interests are often connected to the Challenger's unique historical and cultural experiences, and that these issues are rare. What separates greedy and limited aims Challengers is often how sensitive the Challenger is to historical and cultural value relative to instrumental value. Challengers with limited aims care a lot about territories of historical and cultural importance and less about all else in absolute terms. The Defender is uncertain about whether the Challenger is the limited or greedy type. However, the Defender knows what issues and territories hold normative value for the Challenger with a moderate degree of confidence.

What the Defender knows generates two different strategic contexts in which the Challenger can use costly actions to engender trust. In one, the Challenger's military actions serve her declared peripheral interests. In the other, the Challenger's costly military actions serve her declared core interests. I still assume the Defender is uncertain about the situation-specific costs of the Challenger's military actions, and over how different Challenger-types value declared core interests.

My goal is to understand how the Challenger's choice to use violent military force affects trust in this complex strategic setting. I have two questions. First, Challengers often want to engender trust when they pursue their core interests. Can they do this? If they can, are they forced to forgo issues of high cultural and historical importance that they so desperately care about? Indeed, we motivated this paper with examples in which the Challenger's violent demands plausibly engender trust. It is one thing to say that a Challenger who fights for a declared core interest will not signal aggressive intentions. But it does not intuitively follow violent revision engenders trust. Second, not all declared core interests are alike. In some cases, core interests primarily serve historical and cultural value. In other cases declared core interests serve both instrumental and normative value. How does the value of different strategic goals vary, and what are the strategic implications when the source of a territory's value is ambiguous?

To gain leverage on these questions, I analyzed a game-theoretic model that embeds my motives framework into a classic trust problem. Specifically, I use the game tree presented in Table 1, and make assump-

---

<sup>16</sup>Whether states are entirely ignorant about each other's history is an empirical question that I address in the empirical section.

Table 2: My assumptions about preference variation and information

	My assumptions	Common Formal Assumptions
	<i>Description of Challenger's motives</i>	
Informal	Limited aims Challenger's value core interest revision a lot and peripheral interest revision a little in expectation. There are few issues consistent with core interests.	Limited aims Challengers value revision less than greedy types in expectation.
Parametric	Figure 1(b)-(c). Challenger's type (Greedy $C_G$ , Limited $C_L$ ) is <b>private</b> . $C_G$ value all issues $M$ . $C_L$ value core $H$ and peripheral $L$ . Where $H > L, M > L$ and possibly $H > M$ . Onset of steps (2), (4) Nature determines if issue is core w/ probability $\lambda < .3$ and peripheral $1 - \lambda$ .	Figure 1(a). Challenger's type (Greedy $C_G$ , Limited $C_L$ ) is <b>private</b> . $C_G$ value all issues $M$ , and $C_L$ value all issues $L$ where $M > L$ .
	<i>When the Challenger is given an opportunity to engender trust the Defender...</i>	
Informal	Does not know how sensitive the Challenger is to many situation-specific attributes.	Does not know how sensitive the Challenger is to many situation-specific attributes.
Parametric	Onset of steps (2), (4) Nature draws $k_1, k_2$ and shows it <b>privately</b> to C.	Onset of steps (2), (4) Nature draws $k_1, k_2$ and shows it <b>privately</b> to C.
Informal	Knows with some degree of confidence whether the Challenger's revisionist action serves a core or peripheral interest.	Does not know whether the Challenger's revisionist actions would serve revision over a core or peripheral interest.
Parametric	Onset of steps (2), (4) Nature reveals <u>publicly</u> whether issue is core or peripheral.	Either values are pre-set at the beginning of the game, or Nature shows them privately to C.

tions about information and beliefs presented in Table 2. The model introduces variation in the Challenger's preferences to match the core interest model presented in Figure 1(b)-(c).

I solve the model for Pure Bayesian Equilibrium and focus on pure strategy semi-separating equilibria. The technical analysis appears in Appendix A. Here I intuitively describe the key results. 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 long-term intentions and not greedy intentions. I define  $\alpha_0$  as the Defender's initial (prior) belief that the Challenger's aims are limited. This prior is exogenously set. 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. These beliefs form as the result of equilibrium behavior.

I want to know how violent revision engenders trust. Therefore the quantity that interests me is  $\alpha_1|r - \alpha_1|nr$ . When  $\alpha_1|r - \alpha_1|nr$  is positive it means that if the Defender observes violent revision he trust the Challenger more than if he observed the Challenger avoid an opportunity for violent revision. If instead,  $\alpha_1|r - \alpha_1|nr$  is negative, it means that if the Defender observes violent revision he trust the Challenger less than if he observed the Challenger avoid an opportunity for violent revision. If  $\alpha_1|r - \alpha_1|nr \approx 0$  it means

that violent revision has no affect on trust.

We assume nature draws the situation-specific features for the first revision opportunity independently from the second. Therefore, we can separately solve for equilibrium strategies under two conditions: when the Challenger's first revision opportunity arises over a declared core or peripheral interest.

For now, we focus on situations in which the Challenger faces an initial opportunity to use violent revision to capture a declared core interest. Our first questions is: can Challengers violently pursue their core interests and engender trust? The model illuminates the following result:

**Result 1: Violent revision engenders trust.** If a core interest issue holds so much historical and cultural value relative to instrumental value that a limited aims Challenger unambiguously values it more than the greedy Challenger, then:

1. the Challenger engenders trust by taking the opportunity to pursue violent revision
2. the Challenger engenders mis-trust by avoiding violent revision when given the opportunity to take it.

Therefore  $\alpha_1|r - \alpha_1|nr > 0$ .

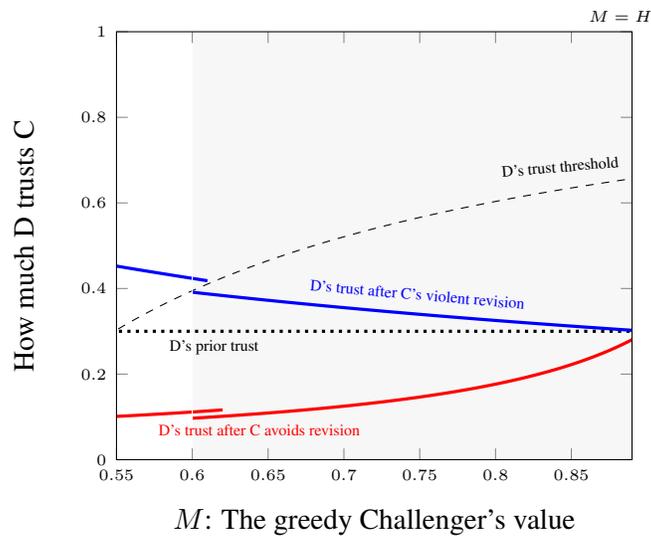
**Result 2.** Result 1 is amplified when the historical and cultural value of revision increasingly and unambiguously exceeds the instrumental value.

Both parts of Result 1 are the opposite of the conventional wisdom from defensive realists and other trust scholars (Kydd 2005; Glaser 2010). Result 1.1 suggests that violent revision engenders trust. Result 1.2 suggests that avoiding violent revision of the status quo can engender mistrust. Why is this the case? Recall that each opportunity for revision is unique. Therefore the Challenger holds private information about the situation-specific cost of revision (this is the cost parameter  $k_1$  in the model). In the semi-separating equilibria a cost-threshold emerges. If the true value of  $k_1$  is below the cost-threshold, the Challenger's benefit from revision outweighs the costs. Therefore, the Challenger selects revision. But if the true cost is above the threshold, the Challenger's cost-benefit analysis drives her to avoid revision. This cost threshold is largely based on the Challenger's value for taking the issue in dispute. Since the Greedy and limited aims Challengers value core interest issues differently, the cost threshold is different. When the limited aims Challenger values her core interests more than the greedy challenger does, the limited aims Challenger's cost threshold for revision is higher than the greedy Challenger's threshold. The reason is simple: the limited aims Challenger intensely values her core interests but the greedy Challenger only values the instrumental aspects of these issues. As a result, the limited aims Challenger is willing to pursue revision at a higher cost level relative to the greedy type.

Since the limited aims Challenger is willing to pursue revision for a wider range of situation-specific costs, violent revision engenders trust. It also follows that avoiding revision engenders mistrust. Using the contrapositive logic, there are fewer situation-specific costs for which the limited aims Challenger prefers to avoid revision.

Figure 2 visualizes the results. The y-axis plots the Defender's level of trust in the semi-separating equilibria. The dotted horizontal line is the Defender's prior trust-level. The blue (red) curve is the Defender's trust after the Defender observes the Challenger use (avoid) violent revision to (not) take a core interest in a semi-separating equilibrium.

Figure 2: Equilibrium Beliefs: Violent Revision Over A Declared Core Interest Engenders Trust



Higher y-axis scores mean that the Defender's trust is higher given equilibrium strategies and what the Defender has observed the Challenger do. Thus, when the distance between the red and blue curves is large the amount of trust following violent revision is large relative to the amount of trust if the Challenger avoids violent revision.

Turning to Result 2, the x-axis varies the Greedy Challenger's value for revision over a declared core interest relative to how much the limited aims Challenger values that issue. To the right ( $M$  is higher) the instrumental value is larger holding the historical and cultural value constant. To the left, the instrumental value is low relative to the normative value of a core interest.<sup>17</sup>

<sup>17</sup>To be clear, varying  $M$  alters the relative value different types hold for core interest issues relative to each other, holding the costs of war constant. We do not derive monotonic predictions if we vary the costs of competition. Once we take into account normative preferences, we can no longer equate predictions from variation in the intensity of preferences to variation in sensitivities to costs.

The shape of the red and blue curves are intuitive given what we now know about the Challenger's cost threshold. Notice that as the greedy and limited aims Challenger's value for a core interest converge, the amount that the Defender learns diminishes. This leads to a third result that helps us understand the implications of ambiguity:

**Result 3: Ambiguity does not lead Defenders to expect the worst.** If a core interest issue is ambiguous in the sense that it involves large amounts of both instrumental and normative value, then violent revision (or not) has no affect on trust. Since it has no affect on trust, violent revision has no impact on strategic competition.

Result 3 is comforting. Past scholars have noted that ambiguity can make learning difficult. They allege that this drives states to expect the worst (Mearsheimer 2001). My result suggests that ambiguity does make learning difficult. But ultimately, it drives states not to alter their beliefs or behavior. As I explain below, this helps explain Britain's long periods of hedging against Hitler's violent revision; among other important cases.

So far, we have focused on how violent revision engenders trust (i.e. a belief). But we have not yet established that the amount of trust that the revision yields impacts the Defender's behavior. We define game-changing trust as a situation where the Defender conditions his competition choice on trust, and trust is contingent on the Challenger's revision.

**Result 4 Violent revision engenders game-changing trust:** Contingent equilibrium exist where, if the Challenger avoids violent revision over a declared core interest, the Defender mistrusts the Challenger so much that he selects competition. However, if the Challenger selects violent revision, the Defender trusts the Challenger enough that he forgoes competition and gives the Challenger future opportunities to revise the status quo. Contingent equilibrium arise when the Defender's prior level of trust ( $\alpha_0$ ) is moderate, the Challenger's situational cost of revision is large (but not too large) in expectation (Expected value of  $k_1$ ), and the instrumental value of revision is unambiguously less than the normative value of revision ( $M$  is low).

The basic insight is that the Defender's competition choice hinges on a trust threshold. This threshold represents the level of risk the Defender is willing to bear given the cost of war, the expectation of victory and the frequency with which a core interest issue is at stake. We plot the trust threshold in Figure 2 using the dashed curve. If the Defender trusts the Challenger a lot ( $\alpha_1$  exceeds the threshold), he prefers to avoid costly competition because he believes that the Challenger is unlikely to capitalize on future opportunities for revision. But if the Defender mistrusts the Challenger ( $\alpha_1$  is less than the threshold), he prefers preventive competition because he believes that the Challenger is likely to pursue revision in the future.

The Challenger's actions engender game-changing trust if they determine whether the Defender's trust

level lies either side of the trust threshold. The unshaded region are the parameter ranges where the Defender conditions his threat of competition on the Challenger's first period revision choice because the Challenger's choice determines if the Defender's trust lies on different sides of the threshold. In these parameter ranges, the Challenger's choice to revise the status quo leads the Defender to trust the Challenger enough to surpass the trust threshold. But the Challenger's choice to avoid revision causes sufficient mistrust that the Defender's beliefs are below the threshold.<sup>18</sup> In the grey region, the Challenger's costly revision does engender trust, but not enough to impact the Defender's action.

### 1.2.1 Connecting the model to past work: When does violent revision engender mistrust?

So far, I have focused on the novel result: violence engenders trust. But in many cases, and also consistent with the conventional wisdom, taking territory can generate mistrust. Can my model also generate this conventional wisdom, and if so, under what conditions?

**Result 5(a): The conventional trust predictions** bear out if the Challenger is faced with a peripheral interest.

This logic for this result is intuitive. Limited aims Challengers care little about their peripheral interests relative to greedy types. When they fight for these issues, they send a strong signal that they care a lot about peripheral interests and, therefore, are likely greedy.

Taking results 1 and 5 together highlights how Defenders exploit information about the Challengers core interests to make opposite inferences from identical actions. When D know that an issue fits C's core interests, then violence can breed trust. When an issue does not fit C's core interests we get the standard logic.

However, past rationalist work does not generate any conditions where violence engenders trust. What is the assumption I change that drives my result? My novel assumption is that limited aims Challengers value core interests as much or more than greedy Challengers. To make clear that this is the necessary assumption, Appendix A.6.1 considers what happens as the type space moves more from the core interest model presented in Figure 1(b) toward the classic assumption presented in 1(a). Formally, I focus on the case if  $H < M$ , and as  $H \rightarrow L$ ?

**Result 5(b) The conventional trust predictions** also bear out if limited aims Challengers value all issues less than greedy types. As  $H \rightarrow L$  the Defender's inferences from observing the Challenger's action over a

<sup>18</sup>The Defender's posterior beliefs ( $\alpha_1$ ) in an equilibrium with game changing trust are dashed.

core or peripheral issue converge.

The standard assumption is that limited aims Challenger's care less about all issues in expectation. Violence only engenders trust if limited aims Challengers value a specific core interests as much or more than greedy Challengers. Trust follows because these high valued issues are relatively rare, and this limits the frequency with which the limited aims Challenger makes revisionist demands.

### **1.2.2 Scope Condition: Variation in the Challenger's motives must create a trust problem.**

As stated, the trust 'problem' assumes that the Defender wants to avoid preventive competition if he is certain that the Challenger holds limited aims. Thus, a trust problem only arises if the limited aims Challenger is sufficiently unlikely to fight in future disputes. For this condition to hold, limited aims Challengers must be sufficiently limited in two important respects (see Appendix A.1.2 for formal statement).

First, C must value peripheral interests sufficiently low given the range of situation-specific costs. Second, C's core interests must be sufficiently scarce. If either condition is violated, then the Defender expects that the limited aims Challenger is highly likely to fight in future disputes. As a result, D selects preventive competition against all Challengers. There is no "trust problem" in this case because the limited aims types are also highly likely to fight in future disputes.

These insights connect my theory to the recent Russo-Ukrainian conflict. At first glance, it might appear that this case works against my theory because Putin invaded a territory that clearly fit within his historical and cultural interests and NATO turned to competition. However, this is a case where Putin's core interest claims were sufficiently vast that NATO was unwilling to tolerate them. At the beginning of the conflict, Putin publicly announced that he held interests to restore Soviet borders. His statement of territorial interests covered Ukraine, Romania, the Baltic states, and Poland. NATO is not willing to tolerate Putin's aims even if they are limited to these historically salient territories.

In fact, my informal insights about how Defenders learn the Challenger's history shed light on one of the most puzzling features of this case. Many have wondered why Putin made expansive territorial claims at the beginning of the conflict. Why didn't he say 'I want Ukraine and nothing else'? My theory suggests that this claim would have been implausible because Russia has a high historical and cultural attachment to many post-Soviet states. By invading Ukraine, Putin made clear he was willing to bear large situation-specific costs to fight for historically salient territories. Russia-experts in NATO would have quickly realized that Ukraine was not the end of Putin's ambition precisely because they know that there are many other territories

of a similar high historical value. In effect, the world's knowledge of Russia's history bounded what Putin could claim.

## 2 Experimental evidence

I generate evidence in support of my mechanism using a survey experiment embedded in a foreign policy scenario. Most experiments randomize state actions. My design is novel because I hold choices constant. I randomly vary the cultural and instrumental value of what states are fighting over. I claim that different historical and cultural details will cause subjects to update in different directions holding a state's invasion choices constant. First I explain my design. Second, I derive explicit hypotheses that connect my theory to the instrument. Third, I report results.

I pre-registered the design, qualtrics print-out, standards for inference, implementation procedure, and R markdown documentation at [REDACTED, see Appendix B for copy].<sup>19</sup> I fielded the experiment on March 25th, 2022. I recruited 200 subjects via lucid online survey platform. Foreign policy vignettes conducted on Lucid have been recently published in prominent political science outlets. Lucid is attractive because it weights recruitment to broadly represents the American population. To account for recent inattentiveness in online responses (Aronow, Kalla, Orr, and Ternovski 2020), I used two pre-treatment attention checks and a minimum time criteria criteria. Per my pre-registration, I analyze the attentive sample (N=182).

Since I theorize historical and cultural details matter, my main study focuses on a real country with rich historical detail. However, I accept that different design choices hold different advantages and threats to validity. Thus, I deployed a second study with an abstract design that I describe in section 2.5.

Many scholars ask what we can learn about theories of elites from public experiments. Recent findings suggest public experiments are worthwhile approximation under several conditions (Kertzer 2020). Consistent with arguments made in Goldfien et al. (2022) a general population experiment is appropriate for my theory because the outcome of interests is beliefs and not policy preferences, and my hypotheses are derived from a rational model (rather individual-level psychological attributes). Beyond the experiment, I use two different experimental designs and historical anecdotes to raise my confidence that my experimental findings plausibly hold real-world impact. Since my theory is about rational inferences drawn from the challenger's observable behaviors, how third-party subjects update is theoretically identical to the beliefs of uninformed

---

<sup>19</sup>These documents also describe how I comply with APSA's principles for engaging human subjects. The study was deemed exempt by my IRB.

defenders observing the same behaviors.<sup>20</sup>

My study focuses on Japan in 2040. Japan has two unique historical features that allowed me to measure my core variables. First, Japan has a history of both expansive (1920s) and limited (1980s) foreign policy ambition. Second, Japan has a different historical and strategic connection to many smaller islands across the East Asian seas. I exploit variation in the real-world context that surrounds three islands.

I code the Amami Islands as a core interest that scores high in historical and cultural importance to Japan but low in instrumental value. Japan discovered the Amami Islands in 625 AD and has controlled them ever since. Currently, the Amami Islands are home to 200,000 Japanese citizens and several cultural sites. However, the Amami Islands have no natural resource deposits and little strategic value for Japan.

I code the Senkaku Islands as a core interest that scores moderate in historical and strategic value. Japan has claimed the Senkaku Islands since 1890 and has a history in fishing around the Islands. Based on this claim, Japan considers the Senkaku Islands a core interest in real life. However, control over the Islands is contested by neighboring states and the Islands are uninhabited. The Senkaku Islands also sit on vast oil reserves.

I code the Liancourt Island as scoring low (but positive) in historical importance and moderate in strategic value. Japan controlled the Liancourt Islands during the 16th Century and again during the Meiji Expansion. But the Liancourt Islands are very close to the Korean coastline and usually administered by Korea.

## 2.1 Vignette Structure

The experiment has three phases, with random assignment occurring only in the third phase. Phase 1 presents factual information about Japan's foreign policy history, existing territorial borders, and Japan's connection to different Islands.

Phase 2 presents plausible information about events in Japan and East Asia between 2022 and 2039. Phase 2 sets up the classic trust problem while providing sufficient detail to account for heterogeneous treatment effects (Dafoe, Zhang, and Caughey 2016). Specifically, I tell subjects that Japan's five main Islands fought a long civil war. While Japan was weak, its neighbors took Japan's smaller Islands. Notably, Korea took the Amami and Senkaku Islands. In 2038 a new regime re-unified Japan and ended the civil war. The new Japanese economy is booming and Japan is becoming the most powerful state in Asia.

---

<sup>20</sup>The prospect of different preferences and motivated reasoning only makes it more difficult to find an effect on beliefs that supports my theory.

Including recently lost territory creates territorial objectives of historical significance. This matches historical cases like China's claim for Taiwan, Hitler's claim for the Rhineland, and Argentina's claim for the Falkland Islands. Including a new regime that is entering a period of strength allows me to encode Japan's history, but creates uncertainty about how sensitive that regime is to issues of history and culture (like [Renshon et al. 2018](#)). To offset alliance dynamics and a dominant China, I tell subjects that the US and China fought a devastating war and that the US has removed its military bases from East Asia.

Phase 2 also presents plausible assessments of Japan's potential foreign policy motives. The final page of the prompt summarizes two potential sets of motives that Japan could hold. These match both my theory and the history of Japan's foreign policy preferences under different regimes. It states,

Japan's foreign policy statement explains that Japan has limited territorial ambitions: "During the 2030s, several Islands of vital historical and cultural significance were stolen from us. We will re-take Islands of historical significance through force if necessary. However, we hold no other territorial ambitions."

A Chinese diplomat disagrees: "Japan's appeal to limited aims is a cover for its true ambition: regional domination. At first, Japan will selectively expand into territories of economic and strategic value to consolidate power. If Japan succeeds at this first step, we won't be strong enough to stop Japan from dominating all of East Asia."

At the end of phase 2, I measure prior (i.e. pre-military action) level of trust using a metric I describe in a moment.

In phase 3, subjects are randomly assigned into one of three crises over different Islands: the Amami, Senkaku or Lincourt Islands. No matter which Island is assigned, the crisis plays out in the exact same way. Korea initiates a plan to populate the Islands with native Koreans and consolidate power over them. Japanese political elites considers two options: invade or not. In every case, Japan invades the Island in dispute.

To be clear, the language I use to describe choices in each treatment arm is identical. I make sure to hold constant Japan's capabilities and expectations of success in battle, the costs Japan will suffer from fighting, Japan's actions (the use of violent force), and Korea's behavior that triggers the crisis in the first place. All that changes is my description of the normative and instrumental value of the three Islands that states are fighting over.

At the end of phase 3, I measure posterior (i.e. post randomly assigned military action) trust.

## 2.2 Measuring Trust

Since my theory, like most theories of trust, focus on two types I use the following measure.

We now want your opinion about how likely these outcomes are relative to each other. In summary, here are the two possibilities:

A. **Limited**: Japan will at most expand its borders to take Islands of historical value.

B. **Expansive**: Japan will opportunistically expand its borders across East Asia. Japan may not always fight. But it will expand when the strategic and economic value outweighs the cost.

How likely are these two outcomes relative to each other?



This measure closely connects to my theory, as opposed to an alternative outcome that may be more subjective and interpreted in different ways by different people.<sup>21</sup>

One concern is that the limited and greedy ambitions I focus on do not capture the main options that subjects consider.<sup>22</sup> To address this, I asked subjects if the two possibilities that I focus on characterize the two most likely possibilities. 89% of respondents believed that these two outcomes were the most likely outcomes.

<sup>21</sup>I forced responses at integer values to prevent fine grained choices that can raise inattentiveness (Lupton and Jacoby 2016).

<sup>22</sup>This would not bias my findings, but would inflate their substantive importance.

## 2.3 Pre-registered Hypotheses

I pre-registered two hypotheses that grapple with two of my most surprising theoretical results. Result 1 states that violent revision over a core interest engenders trust. I claim this result is most likely to obtain when the issue in dispute scores high on historical and cultural value and low on instrumental value. Thus, my first hypotheses concerns how subjects who observe the Amami Island dispute update over time:

**Hypothesis 1:** Subjects that observe Japan use violent military force to re-take the Amami Islands will increase their trust in Japan. (Posterior trust is greater than prior trust).

Existing studies posit several reasons that I should not observe this result. Recall the conventional prediction is the opposite of what I expect. Normative scholars intuit that a subject can appreciate historical context in this case (e.g [Mitzen 2006](#); [Hopf 1991](#)). However, they theorize that subjects, at most, do not grow mistrustful in a case like this. Thus, they would expect that violent revision engendered trust. As I showed in result 5, this no-affect result would also hold if greedy Challengers were as sensitive to historical value as limited aims Challengers.<sup>23</sup>

Result 2 states that to the extent that an issue serves historical and cultural value will mediate trust. Interpreting this result in real life requires a more detailed appreciation of the determinants of core and peripheral interests. Drawing from a diverse literature about the determinants of state preferences, I argued that limited aims states commonly seek out territories of historical and cultural value (e.g [Carter and Goemans 2011](#)), and greedy states seek out territory of instrumental value. As stated, my focus on historical value was an approximation. However, if the approximation is reasonable and broadly held, I will be able to validate the following hypothesis:

**Hypothesis 2:** Subjects that observe Japan use violent military force to re-take the Amami Islands will trust that Japan's long run intentions are limited more than subjects who observe Japan use violent military force to take the Liancourt Islands.

In an easier test that is consistent with result 2, 3, and 5, I could have chosen a comparison territory that had no historical attachment to Japan (e.g. a Fijian Island). I used the Liancourt Islands as the comparison group because it is a tough test for my theory that takes the logic of pre-text seriously ([Press 2007](#)). To stack the deck in favor of pre-text ruining my result, I told subjects that Japan controlled the Liancourt Island in the 16th Century and had a historical claim to it. After subjects observe Japan fight for the Liancourt Island, they read a statement from the Japanese foreign minister that justifies the invasion in terms of Japan's historical

---

<sup>23</sup>I address this assumption explicitly in the next experiment.

context.<sup>24</sup> If pre-text is easy to generate, then subjects will respond the same way to Japan using force to take the Amami and Liancourt Islands. As a result, I should see no difference between these two treatment arms.

My theory also grappled with the strategic implications when issues in dispute were ambiguous because they served both instrumental and historical value. Others suggest that ambiguity drives states to expect the worst. Result 3 shows that when issues ambiguously serve both instrumental and normative value, then violence has no affect on trust.

**Conjecture:** Observing Japan use violent military force to re-take the Senkaku Islands will not affect a subject's level of trust over time.

I pre-registered this as a speculative conjecture rather than a hypothesis for two reasons. First, the inference that subjects draw depends on (a) the amount of normative and instrumental value involved in the issue in dispute; and (b) how sensitive subjects believe that limited and greedy Challengers are to these different values. I could not precisely measure perceptions of the Challenger's sensitivity to different sources of value.

Second, this test also verifies the real-world domain to which my theory applies. If I find evidence that fighting over the Senkaku Islands generates trust, it would illustrate the far-reaching implications of my theory. The reason is that even when territories have huge economic value, subjects interpret the normative value as the dominant component that drives revisionist actions. If instead, I find that violence engenders mis-trust it would suggest that the conventional wisdom dominates in many real-world cases.

Based on my prior belief that hypothesis 1 would be hard to validate, my power analysis suggested that I needed more subjects in the Amami Islands treatment arm to find an affect. Thus, I randomly assigned subjects to treatment arms using an i.i.d random variable ratio of 2:1:1 (Amami:Senkaku:Liancourt).<sup>25</sup>

## 2.4 Results

For each hypothesis I pre-registered a permutation tests to determine the difference in group means. Based on pre-registered definitions of support thresholds, I find support for hypothesis 1<sup>26</sup> and 2.<sup>27</sup> Appendix B.4 reports R-markdown print-out for the analytical procedure and relevant data tables to generate figures.

---

<sup>24</sup>To be clear, this justification appears in all treatment groups.

<sup>25</sup>This ratio made it harder to support H2 but easier to support my conjecture (which predicts no affect). As we shall see, the mean respondent for the conjecture inferred trust (but not at a significant level).

<sup>26</sup>Means are distinguishable at  $p = .023$ .

<sup>27</sup>Means are distinguishable at  $p = 0.011$ .

I visualize the results in Figure 3. Figure 3(a) plots the sample computed mean and 95% confidence intervals for subjects trust levels. Recall the scale labels -10 definitely limited and 10 as definitely expansive. The post-crisis trust levels are in blue. The pooled prior estimate is in red.<sup>28</sup> These summaries closely correspond with the analytical test, and therefore provide the easiest visual presentation to validate my two hypothesis.

Hypothesis 1 focus on the Amami Islands treatment and asks: do subjects who observe violent revision grow more trusting? As Figure 3(a) shows, subjects were more confident that Japan held limited rather than greedy aims after they observed Japan use force to take the Amami Islands. The difference between the prior and the posterior is large.<sup>29</sup> In contrast to the conventional wisdom, this result suggests that violent revision can engender trust when Challengers are using force to take an unambiguous core interest.

Hypothesis 2 contrasts the post-crisis beliefs of subjects that observed the Amami and Liancourt Island treatments. Figure 3(a) shows a clear statistical difference in how much subjects respond to these two crisis scenarios. In the Amami Islands case they grew trustful. In the Liancourt Islands case they grew mistrustful. This confirms that subject beliefs are mediated through cultural context holding all other features of the scenario constant. Fighting over clear core interests engender trust, but fighting over issues with only a small amount of cultural attachment engenders statistically significantly less trust.

Finally, the Senkaku Island results help us explore the bounds of violent trust. The Senkaku Islands involved moderate amounts of historical and strategic value. Consistent with the conjecture, I could not distinguish the Senkaku Islands from its prior at pre-registered levels of confidence.<sup>30</sup> However, the average subject did grow more trusting. This suggests that the empirical domain of violent trust could be quite large. At minimum, the Senkaku Island's large oil deposits did not instantly drive subjects to expect the worst.

Figure 3(b) plots how much subjects change over time (I subtract their post-crisis trust from their pre-crisis trust). Theoretically, a subject could score a maximum of 20 (minimum of -20) if they shifted from definitely limited (expansive) to definitely expansive (limited). The black line at 0 represents subjects that did not alter their beliefs. Shaded in red is the density of subjects that grew mis-trustful. Shaded in grey are subjects who grew trustful.

The distributions of updated are consistent with my theory. The distribution of subjects that observed

---

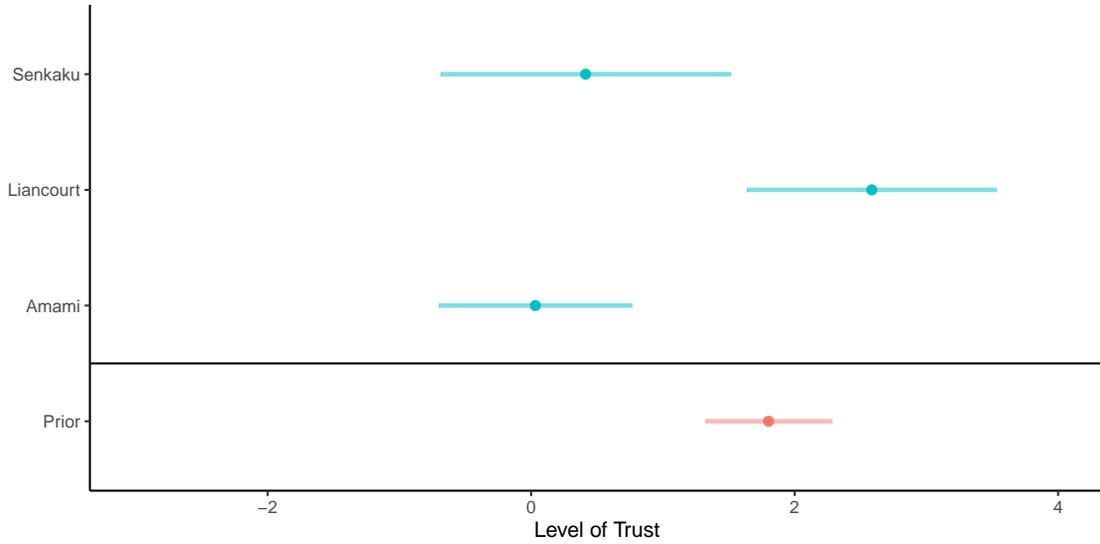
<sup>28</sup>We plot the pooled prior for ease. It does not alter our substantive interpretation. The analytical tests and density plots examine change over time.

<sup>29</sup>I visualize the pooled prior for ease. However, the analytical test contrasts the prior with the Amami group and renders a strong statistically significant difference.

<sup>30</sup>I was close. With  $p = .22$  violent revision over the Senkaku Island engendered trust.

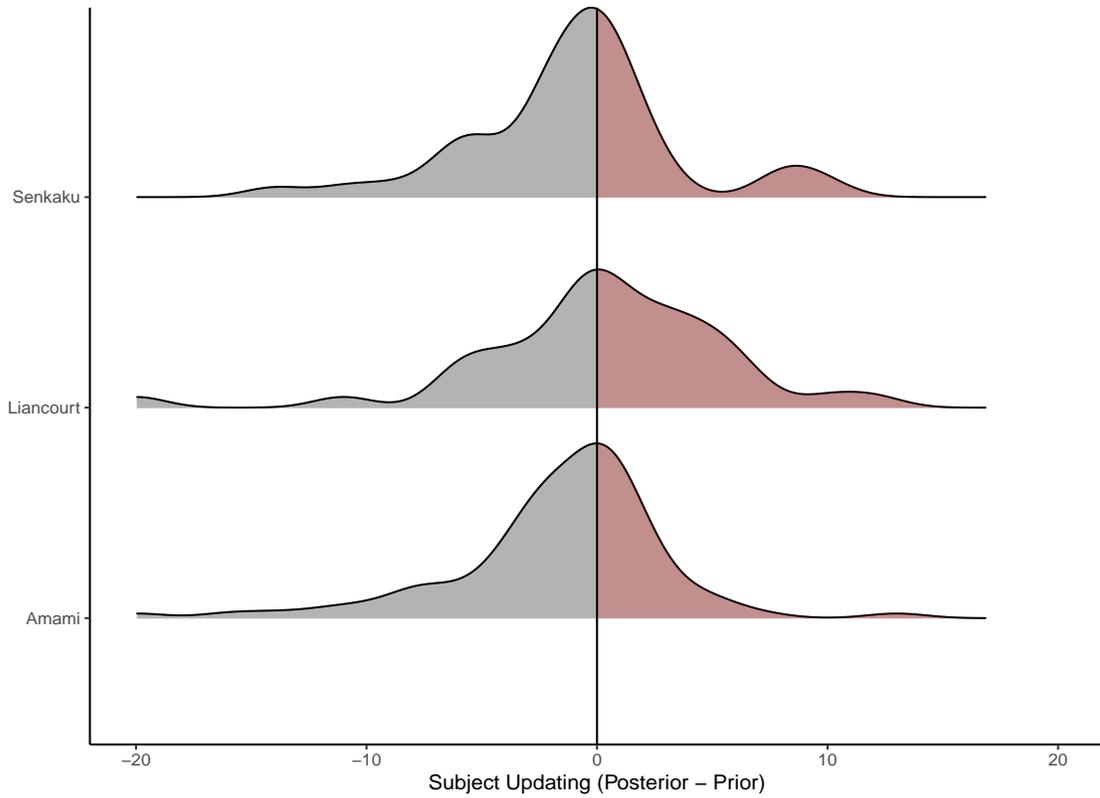
Figure 3: Experimental Results

(a) Mean and 95% Confidence of Trust Answers



Summary of absolute trust scores after observing a randomly assigned violent invasion (blue) and the pooled prior belief (red).  
Larger numbers indicate lower levels of trust.

(b) How individuals change in response to a crisis.



The black line represents no change. Red (grey) regions are subjects that became more mistrustful (trustful) following a military intervention.

the Amami Crisis is left skewed with a relatively long tail towards growing trustful. In contrast, the mass of subjects that observed the Liancourt crisis grew slightly mistrustful (right skew). There is a longer tail here towards mistrust. To be clear, subjects update in both directions in all treatment groups. The fact that violence can increase trust in any treatment group is surprising for the existing literature. In the context of my theory the different directions of updating make sense if we assume that individual analysts can vary in their interpretations in their assessment of how sensitive limited aims Japan is to very old cultural claims (we will validate this point with textual responses below). Despite this individual-level variation, the overall shape of the distributions, and the average response is consistent with my theory.<sup>31</sup>

#### **2.4.1 The Mechanisms Described in Text Responses**

My causal mechanism suggests that Defenders realize that limited aims Challengers intensely value history and culture. Therefore, they will fight for issues of high historical and cultural importance even if it involves considerable military costs or other situation-specific costs. In contrast, greedy Challengers will avoid revision because they are less sensitive to unique historical benefits and still care about the situation-specific costs. To see if this mechanism impacted assessments I asked “Did Japan’s behavior during this military crisis influence your assessment? If so, why?”

Of the subjects that provided a reason that the crisis altered their beliefs,<sup>32</sup> 2/3 detailed a logic that was broadly consistent with my theory. Here are some examples of strong support. One subject assigned to the Amami Islands treatment stated, “Yes, because it showed that Japan valued the historical importance of Amami over the little resources they had there.” Another stated, “Japan took islands of little strategic value to restore historical territory. This act leads me to believe more firmly that Japan is expanding only to recover its historic territory rather than economic and strategic expansion into territory not historically theirs.”

Another stated, “Yes because if they invaded, the likelihood of it having negative consequences for Japan risk would outweigh the gains” Notably, this captures my exact logic in that only those sensitive to normative value are willing to bear these costs.

By contrast, there were many more mixed responses among those assigned to the Liancourt Island

---

<sup>31</sup>Notice that several responses have extreme scores (e.g. -20, 20) that seem implausible. I conjecture this is a result of inattentiveness. I included a post-treatment attention check that asks if Japan invaded? Remove subjects who fail this check leaves N=163 and omits the most extreme responses. My analytical results are stronger and the visualizations are more compelling if I remove these subjects.

<sup>32</sup>Many wrote NA, or left the question blank. Some others did not provide a logic for their beliefs.

treatment. The majority (3/4) who detailed a response described a logic of mistrust that is consistent with my theory. One stated that “Because it was only limited historical significance to them. They were willing to have casualties for something of not a big significance, thereby suggesting they might expand to others as well.” Even though the vignette does not include the word pre-text, another noted that “The crisis was just the pretext,” to explain why she raised her belief that Japan was greedy. Another said, “The decision to expand beyond the 2 islands they originally had was based on their vision of maintaining East Asian dominance. It appeared or perhaps implied that, the ‘spirit’ of reclaiming what was lost went beyond those intentions to risky self centered intentions.” One subject saw this crisis grew more mistrustful (by 3 points) “because I realize that there isnt enough information to determine what ‘culturally significant’ means and how many islands it might pertain to.”

A minority (1/4) suggested that Japan’s limited historical connection led to small amounts of trust. One stated, “Same outlook as before but this information added the additional HISTORICAL acquisition. So while it marginally influenced my opinion,it didn’t CHANGE it.” However, no response was as strong in terms of trust-building when compared to the Amami descriptions above.

## 2.5 Second (abstract) study

After my main study was complete, I fielded a second study. It confirms that the support for H1 and H2 I found in my main study was not an artifact of my design choices; and tests two additional implications. To address recent debates about abstraction and realism (see [Brutger, Kertzer, Renshon, Tingley, and Weiss 2022](#)) my second study retains the same structure as the first. However, it replaces Japan with a fictional country called Bandaria. It also eliminates all non-essential detail including many historical facts, diplomatic messages, and crisis details. Appendix C reports the vignette. Appendix C.2 reports the R markdown documentation for analytical procedures and relevant data tables to generate figures.

The study includes two simplified variants of the two treatment arms from the main study. In both arms, Bandaria is always faced with a crisis over the Davenport Island and always fights. I randomly vary whether the Davenport Island holds (a) high historical but no strategic value; (b) no historical but high strategic value.<sup>33</sup> I exploit these treatment arms to re-test hypothesis 1 and 2.

To test an additional implication, I replaced the ambiguous treatment arm (Senkaku) with a ‘no fight’ treatment arm. Identical to arm (a), the crisis erupts over an issue of high historical importance. But unlike

---

<sup>33</sup>Because the Island is hypothetical I can hold it constant and vary how I describe it.

arm (a), I tell subjects that Bandaria chooses not to revise the status quo through force. This allows me to test one additional hypothesis:

**Hypothesis 3:** Subjects that observe Bandaria use violent military force to fight for an Island of high historical value will trust that Bandaria's long run intentions are limited more than subjects who observe Bandaria not fight.

Based on the main experiment, and what I know about stronger affects in abstract studies, I recruited 100 subjects and distributed them evenly across treatment groups. I use the exact same R code and standards of inference from the main experiment.<sup>34</sup> Figure 4 visualizes the results of treatment groups in the same format as Figure 3, and annotates the results of H2 and H3.

The analytical tests find support for H1 ( $p=.016$ )<sup>35</sup> and H2 ( $p < 0.001$ ). This raises my confidence that my original findings were not a function of the historical or diplomatic details in my main design.

An analytical test also strongly supports H3 ( $p = .011$ ): subjects who observed Bandaria fight in a historically salient crisis were more trusting than subjects who observed Bandaria not fight. To my knowledge, and consistent with my novel mechanism, this is the first experiment that shows avoiding violent territorial revision creates mistrust more than violent revision.

Although I did not pre-register it, the Figure shows that there is no statistical difference in trust for those who observed (a) Bandaria fight for a peripheral interest, and (b) Bandaria avoid fighting for a core interest. Given that both are different from the case in which Bandaria fights for a core interest, it suggests that historical context has powerful affects in how it mediates inferences from strategic choices.

A closer look at text responses lends additional support to my mechanism. As in the main study, I asked, "Did the crisis alter your assessment? If so, why?" One subject who observed Bandaria avoid violent revision over a historical issue stated, "Yes, it proved that Bandaria is not really that interested in land of historical value because, if it was, it would have fought to gain this land." Another noted "Yes, they had opportunity to seize for historic reasons and chose not to," another noted, "Their choice to not invade the island of Davenport suggests a focus on economic might." There are several others.

Finally, and as discussed in section 1.1, my novel results assume that onlookers intrinsically believe that limited aims Challengers can value their core interests *as much* or *more than* greedy challengers. If this was not true, I would not have found support for my hypotheses. To lend direct support to this assumption, I asked subjects post-survey if they agreed with the following statement: 'States with limited aims care more

---

<sup>34</sup>H3 follows H2s estimation.

<sup>35</sup>H1 is omitted from the Figure for visual ease.

about issues of historical and cultural importance than states with expansive interests.’ As Figure 5 shows, 59% of subjects agreed (i.e. ‘more than’), and another 28.5% had no opinion (i.e ‘as much’). Only 11.5% of subjects disagreed.

### 3 Anecdotes and puzzling quantitative results

I now illustrate how my theory fits with existing data. The cases I discuss are all complex, and multi-causal. I do not provide a detailed account of any of them. Rather, I use my framework to (a) illuminate features of critical cases that historically minded scholars find difficult to explain; (b) provide a logic to reconcile different historical accounts within critical cases; and (c) show that the broader framework of motives and trust could extend to many different trust scenarios.

My theory is consistent with 4 puzzling episodes of trust during Anglo-German relations in the 1930s. Archival researchers have asked why British elites chose to appease Hitler despite Hitler’s repeated use of force to take territory. Although the details vary, scholars identify crises — for example, the re-militarization of the Rhineland and the Anschluss — as key events that should have led the British to conclude that Hitler had vast aims. Others identify how Hitler violated arms control agreements. Since the British did not update their beliefs, these studies conclude that either British elites were irrational, or naive (see [Yarhi-Milo 2014](#); [Gilbert 1972](#); [Barnett 1986](#)).<sup>36</sup>

Yet British assessments are well explained by mapping Hitler’s revision onto his declared core interests. The British believed that Hitler may have been motivated by uniting ethnic-Germans under one government. Each of these crises is consistent with that basic principle. For example, when Hitler re-militarized the Rhineland, the British War Secretary told the German Ambassador, “through the British people were prepared to fight for France in the event of a German incursion into French territory, they would not resort to arms on account of the recent occupation of the Rhineland. The people did not know much about the demilitarization provisions and most of them probably took the view that they did not care ‘two hoots’ about the Germans reoccupying their own territory.”<sup>37</sup>

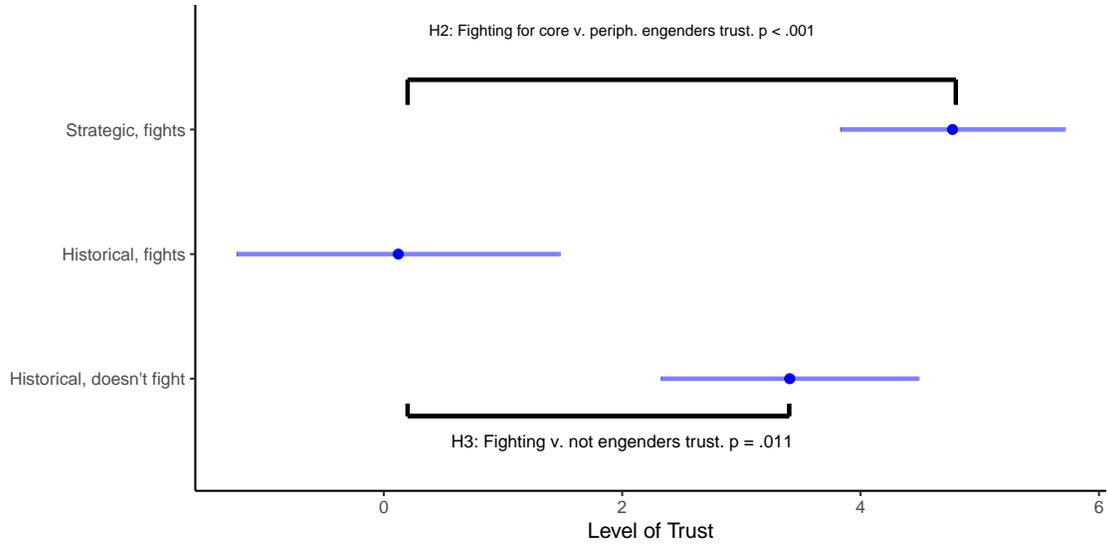
The shift in British assessments following the failure of the Munich Agreement Agreement are also consistent with my theory. In 1934, Hitler mobilized forces to take Czechoslovakia. Alarmed by these

---

<sup>36</sup>One exception is [Ripsman and Levy \(2008\)](#). Their account is consistent with mine. As they point out, the most senior British elites were, at best, uncertain about Hitler’s strategic intentions up until Hitler violated Munich. They do not offer an explanation for why British trust is invariant.

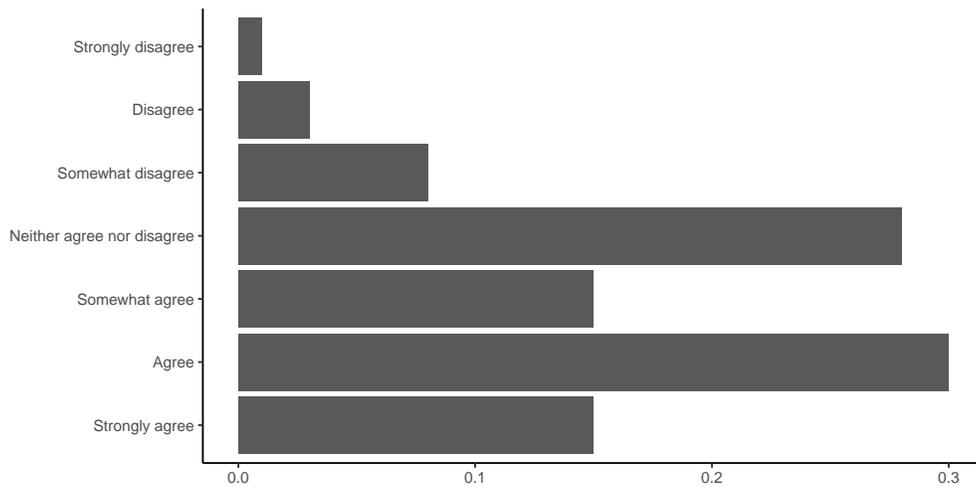
<sup>37</sup>Quoted in [Weinberg \(1980\)](#) p259.

Figure 4: Results of Abstract Experiment



Mean and 95% confidence intervals of absolute trust scores for three treatment groups. Larger numbers (to right) indicate lower levels of trust.

Figure 5: States with limited aims care more about issues of historical and cultural importance than states with expansive interests (Percentages).



events, the British attempted to mediate a settlement. Ultimately, Hitler accepted territorial concessions over only the German speaking parts of Czechoslovakia. based on this compromise, the British were optimistic about peace in Europe. It was only after Hitler took the non-German speaking parts of Czechoslovakia that the British prime minister became convinced that Hitler's aims were expansive.<sup>38</sup> The failure of the Munich Agreement was not the first time that Hitler had taken territory, threatened force, or even violated an international agreement. It was, however, the first time that Hitler took territory that was inconsistent with his declared core interests.

Several other cases plausibly fit my theory that either violence over core interests engenders trust, or has no affect at all. One example is British assessments that Prussian aims were limited following the Danish-Prussian wars and the Franco-Prussian wars.<sup>39</sup> Others are British assessments of the United States circa 1900, and American assessments of the Soviet Union circa 1940. Another is American assessments of Argentina following the Falkland Islands Dispute.

My theory also helps clarify experimental results that cut against the conventional wisdom. For example, [Gottfried and Trager \(2016\)](#) report a survey experiment based on an ongoing Russo-American dispute over oil-rich arctic territories. They find that Russia's "aggressive foreign rhetoric makes it harder for a [American] leader to offer a dramatic settlement involving substantial concessions to preserve the peace," and also "dramatically increases the approval of presidents who prosecute successful wars." This puzzles scholars because the US should be willing to make concessions over these issues when Russia shows resolve, given that they are peripheral to American core interests ([McManus 2014](#)). Consistent with my theory, [Gottfried and Trager \(2016\)](#) tell participants that Russian interests in the arctic are not related to Russia's core interests. Therefore, subjects that observe Russia threaten force to take this territory, observe Russia threaten a peripheral interest. Consistent with my theory, when Russia fights for a peripheral interest subjects are more likely to support the use of force against Russia at the first available chance and are less likely to support short-term concessions.

## 4 Conclusion

Scholars from a broad range of traditions believe that trust is essential for avoiding catastrophic conflict ([Weisiger 2013](#); [Glaser 2010](#); [Mitzen 2006](#)). Against the conventional wisdom, I argued that when states

---

<sup>38</sup>This account is consistent with ([Ripsman and Levy 2008](#); [Bullock 1971](#); [Hillgruber 1974](#))

<sup>39</sup>It is also consistent with wide-spread concern that followed once Prussia made demands for French-speaking France.

contest an issue with enormous historical and cultural value they can engender trust through costly militarization, violating arms control agreements, or even a military invasion to take territory through force. In contrast, states that avoid violent behavior signal mistrust. I support this theoretical claim with a survey experiment that shows a revisionist Japan in 2040 can engender trust by using violent military force to invade the Amami Islands. In an abstract experiment I show that a fictional belligerent also engenders mistrust by avoiding violent territorial revision.

I also argue that ambiguity over the source of value (i.e. territories that hold historical, cultural, and instrumental value) can make learning more difficult. However, I theorized that this does not drive states to expect the worst ([Mearsheimer 2001](#)). Rather, it leads to no change in beliefs, and this can lead Defenders to continue cooperative policies even as the Challenger pursues violent revision. I supported this claim by showing experimental subjects did not engender mistrust when revisionist Japan invaded the Senkaku Islands. I also illustrated it with several historical anecdotes including British trust following Hitler's violent revision in the Rhineland and Austria.

Taken together, these results show that the same violent action drives opposite updating on questions of trust depending on the historical context that surrounds the action. Thus, the relationship between military choices, trust, and great power competition is vastly more complicated than rationalist scholars realize. 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 reverse 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. It suggests that empirical models will enhance predictive accuracy if they combine data on issue-specific disputes ([McLaughlin Mitchell and Hensel 2007](#)) with data on the historical and cultural nature of claims ([Abramson and Carter 2016](#); [Schultz 2017](#)) to appreciate core and peripheral interests.

The basic logic of intense cultural preferences likely impacts beliefs in other important information problems. I conjecture that cultural context impacts how the reputation for resolve forms ([Kertzer 2016](#)). It is plausible that avoiding force over a peripheral interest could communicate high resolve to fight for core interests in the future. Therefore it is possible to construct settings where avoiding crisis entry can signal high resolve to fight in future disputes. By focusing on variation over the issues in dispute we can start to appreciate why leaders with different psychological traits, and experiences, or who are sensitive to the nor-

mative preferences of their constituents, may value some issues more than others. This has important policy implications. It is plausible that the US could shirk on its NATO alliances and strengthen its commitment to defend Japan, because letting Europe fall could communicate a pivot to Asia. In these cases, both strategy and cultural context have a precise interaction that needs closer scholarly attention. The outcome of this interaction will depend on how others understand the United States core interests.

This framework can help bridge the gap between theory and state practice. After interviewing 200 national security professional for a broader project on how the United States makes inferences about China's intentions, I have learned that intelligence analysts deeply appreciate how strategic incentives to misrepresent and cultural context intersect. For example, China analysts interpret China's violent actions over Taiwan and the South China Sea differently because the historical context is different. Therefore, they draw a difference inference about what could motivate China from seemingly similar actions. Thus, even though history can make theory complicated, it seems critical for explaining critical episodes in great power competition.

## References

- Abramson, Scott F. and David B. Carter, 2016. The Historical Origins of Territorial Disputes, *American Political Science Review*, **110**(4), 675–698.
- Aronow, Peter, Joshua Kalla, Lilla Orr, and John Ternovski, 2020. Evidence of Rising Rates of Inattentiveness on Lucid in 2020., Tech. rep., SocArXiv.
- Axelrod, Robert, 1980. More Effective Choice in the Prisoner's Dilemma, *Journal of Conflict Resolution*, **24**(3), 379–403.
- Axelrod, R. and R. Iliev, 2014. Timing of cyber conflict, *Proceedings of the National Academy of Sciences*, **111**(4), 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**(1), 5–36.
- Bas, Muhammet A and Andrew J Coe, 2016. A Dynamic Theory of Nuclear Proliferation and Preventive War, *International Organization*, **70**(4), 655.
- Braumoeller, Bear F, 2008. Systemic Politics and the Origins of Great Power Conflict, *American Political Science Review*, **102**(1), 77–93.
- Brutger, Ryan, Joshua Kertzer, Jonathan Renshon, Dustin Tingley, and Chagai Weiss, 2022. Abstraction and Detail in Experimental Design. American Journal of Political Science, *American Journal of Political Science*.
- Bullock, Alan, 1971. Hitler and the Origins of the Second World War, in *The Origins of the Second World War*, pp. 189–224, Macmillan Education UK, London.
- Carr, Edward Hallett, 1945. *The Twenty Years' Crisis, 1919-1939.*, Harper Row.
- Carter, David B. and Hein E. Goemans, 2011. The Making of the Territorial Order: New Borders and the Emergence of Interstate Conflict, *International Organization*, **65**(02), 275–309.
- Coe, Andrew and Jane Vaynman, 2019. Why Arms Control Is So Rare, *American Political Science Review*, pp. 1–14.
- Coe, Andrew J. and Jonathan N. Markowitz, 2021. Crude Calculations: Productivity and the Profitability of Conquest, *International Organization*, **75**(4), 1058–1086.
- Crescenzi, Mark J. C., 2007. Reputation and Interstate Conflict, *American Journal of Political Science*, **51**(2), 382–396.
- Dafoe, Allan, Jonathan Renshon, and Paul Huth, 2014. Reputation and Status as Motives for War, *Annual Review of Political Science*, **17**(1), 371–393.
- Dafoe, Allan, Baobao Zhang, and Devin Caughey, 2016. Confounding in Survey Experiments.
- Debs, Alexandre and Nuno P. Monteiro, 2014. Known Unknowns: Power Shifts, Uncertainty, and War, *International Organization*, **68**(01), 1–31.

- Edelstein, David M., 2019. *Over the horizon : time, uncertainty, and the rise of great powers*, Cornell University Press.
- Fearon, James D. and Alexander Wendt, 2002. Rationalism v. Constructivism: A Skeptical View, in *Handbook of International Relations*, pp. 52–72, SAGE Publications Ltd, 1 Oliver’s Yard, 55 City Road, London EC1Y 1SP United Kingdom.
- Finnemore, Martha, 1996a. *National Interests in International Society*, Cornell studies in political economy, Cornell University Press, Ithaca, N.Y.
- Finnemore, Martha, 1996b. Norms, culture, and world politics: insights from sociology’s institutionalism, *International Organization*, **50**(02), 325.
- Frieden, Jeffrey a., 1999. Actors and Preferences in International Relations, in *Strategic Choice and International Relations*.
- Gilbert, Martin, 1972. *The Roots of Appeasement*, Weidenfeld and Nicolson, London.
- Glaser, Charles and Chaim Kaufmann, 1998. Offense Defence Balance and Can We Measure It, *International Security*, **22**(4), 44–82.
- Glaser, Charles L., 1995. Realists as Optimists: Cooperation as Self-Help, *International Security*, **19**(3), 50–90.
- Glaser, Charles L., 2010. *Rational Theory of International Politics*, Princeton University Press, Princeton.
- Glaser, Charles L., 2015. A U.S.-China Grand Bargain? The Hard Choice between Military Competition and Accommodation, *International Security*, **39**(4), 49–90.
- Goemans, Hein E. and Kenneth A. Schultz, 2017. The Politics of Territorial Claims: A Geospatial Approach Applied to Africa, *International Organization*, **71**(1), 31–64.
- Gottfried, Matthew S. and Robert F. Trager, 2016. A Preference for War: How Fairness and Rhetoric Influence Leadership Incentives in Crises, *International Studies Quarterly*, **60**(2), 243–257.
- Gurantz, Ron and Alexander V. Hirsch, 2017. Fear, Appeasement, and the Effectiveness of Deterrence, *The Journal of Politics*, **79**(3), 1041–1056.
- Haynes, Kyle and Brandon K. Yoder, 2020. Offsetting Uncertainty: Reassurance with Two-Sided Incomplete Information, *American Journal of Political Science*, **64**(1), 38–51.
- Hillgruber, Andreas, 1974. England’s Place in Hitler’s Plans for World Domination, *Journal of Contemporary History*, **9**(1), 5–72.
- Hopf, Ted, 1991. Polarity, the Offense-Defense Balance, and War, *American Political Science Review*, **85**(2), 475–493.
- 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**(02), 167–214.

- Joseph, Michael F., 2020. A Little Bit of Cheap-Talk is a Dangerous Thing: States Can Communicate Intentions Persuasively and Raise the Risk of War in the Processes., *The Journal of Politics*, p. 709145.
- Katzenstein, Peter J., 1996. *The Culture of National Security*, Columbia University Press, NY.
- Kertzer, Joshua D., 2016. *Resolve in international politics*, Princeton University Press.
- Kertzer, Joshua D., 2017. Resolve, Time, and Risk, *International Organization*, **71**(S1), S109–S136.
- Kertzer, Joshua D., 2020. Re-Assessing Elite-Public Gaps in Political Behavior, *American Journal of Political Science*.
- Kertzer, Joshua D. and Kathleen M. McGraw, 2012. Folk Realism: Testing the Microfoundations of Realism in Ordinary Citizens, *International Studies Quarterly*, **56**(2), 245–258.
- Kydd, Andrew H., 2005. *Trust and Mistrust in International Relations*, Princeton University Press, Princeton, N.J.; Woodstock.
- Little, Andrew T. and Thomas Zeitzoff, 2017. A Bargaining Theory of Conflict with Evolutionary Preferences, *International Organization*, **71**(3), 523–557.
- Lowenthal, Mark M., 2019. *Intelligence : from secrets to policy*.
- Lupton, Robert and William G. Jacoby, 2016. The Reliability of the ANES Feeling Thermometers: An Optimistic Assessment, in *Annual Meetings of the Southern Political Science Association*.
- McGillivray, Fiona and Alastair Smith, 2008. *Punishing the prince : a theory of interstate relations, political institutions, and leader change*, Princeton University Press.
- McLaughlin Mitchell, Sara and Paul R. Hensel, 2007. International Institutions and Compliance with Agreements, *American Journal of Political Science*, **51**(4), 721–737.
- McManus, Roseanne W, 2014. Fighting words, *Journal of Peace Research*, **51**(6), 726–740.
- Mearsheimer, John J, 2001. *The Tragedy of Great Power Politics*, Norton, New York.
- Mercer, Johnathan, 1996. *Reputation and International Politics*, Cornell University Press.
- Mitzen, Jennifer, 2006. Ontological Security in World Politics: State Identity and the Security Dilemma, *European Journal of International Relations*, **12**(3), 341–370.
- Moravcsik, Andrew, 1997. Taking Preferences Seriously: A Liberal Theory of International Politics: Erratum, *International Organization*, **52**(4), 229.
- O’Neill, Barry, 1999. *Honor, Symbols, and War*, University of Michigan Press, Ann Arbor, MI.
- Powell, Robert, 1996. Uncertainty, Shifting Power, and Appeasement, *The American Political Science Review*, **90**(4), 749–764.
- Powell, Robert, 2006. War as a Commitment Problem, *International Organization*, **60**(01).
- Powers, Kathleen E., 2022. *Nationalisms in international politics*, Princeton University Press.
- Press, Daryl G., 2007. *Calculating Credibility: How Leaders Assess Military Threats*, Cornell University Press.

- Qimao, Chen, 1996. The Taiwan Strait Crisis: Its Crux and Solutions, *Asian Survey*, **36**(11), 1055–1066.
- Ramsay, Kristopher W., 2017. Information, Uncertainty, and War, *Annual Review of Political Science*, **20**(1), 505–527.
- Renshon, Jonathan, 2016. Status Deficits and War, *International Organization*, **70**(03), 513–550.
- Renshon, Jonathan, Allan Dafoe, and Paul Huth, 2018. Leader Influence and Reputation Formation in World Politics, *American Journal of Political Science*.
- Ripsman, Norrin M. and Jack S. Levy, 2008. Wishful Thinking or Buying Time? The Logic of British Appeasement in the 1930s, *International Security*, **33**(2), 148–181.
- Schultz, Kenneth A., 2017. Mapping Interstate Territorial Conflict, *Journal of Conflict Resolution*, **61**(7), 1565–1590.
- Slantchev, Branislav L., 2010. Feigning Weakness, *International Organization*, **64**(03), 357–388.
- Snyder, Jack, 1993. *Myths of Empire: Domestic Politics and International Ambition*, Cornell University Press, Ithaca, N.Y.
- Svolik, Milan, 2006. Lies, Defection, and the Pattern of International Cooperation, *American Journal of Political Science*, **50**(4), 909–925.
- Trager, Robert F., 2010. Diplomatic Calculus in Anarchy: How Communication Matters, *American Political Science Review*, **104**(02), 347–368.
- Trager, Robert F., 2016. The Diplomacy of War and Peace, *Annual Review of Political Science*, **19**(1), 205–228.
- 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**(3), 526–545.
- Vaynman, Jane, 2021. Domestic Political Shifts and the Design of Security Agreements.
- Waltz, Kenneth N., 1979. *Theory of international politics*, McGraw-Hill, Reading, Mass.
- 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.
- Weisiger, Alex and Keren Yarhi-Milo, 2015. Revisiting Reputation: How Past Actions Matter in International Politics, *International Organization*, **69**(2), 473–495.
- Wendt, Alexander, 1992. Anarchy is what States Make of it, *International Organization*, **46**(2), 391–425.
- 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**(1), 130–145.

# The Appendix

## Appendix

### Table of Contents

---

<b>A Formal Presentation of theory</b>	<b>39</b>
A.1 Set up . . . . .	39
A.2 First issue is a peripheral interest . . . . .	41
A.3 Violent revision engenders trust when faced with a core interest crisis . . . . .	44
A.4 Generating the Comparative static plot in the Manuscript . . . . .	48
A.5 C's decision to fight engenders game changing trust: . . . . .	49
A.6 Exploring the Boundaries of violent trust. . . . .	50
A.7 Defenders are uncertain if a territory is a genuine core/peripheral interests. . . . .	51
<b>B Main (Japan) experiment</b>	<b>53</b>
B.1 Pre-analysis Plan . . . . .	53
B.2 Qualtrics print out . . . . .	63
B.3 Pre-registered analytical procedure in R . . . . .	83
B.4 R markdown for analysis of results . . . . .	89
<b>C Abstract experiment</b>	<b>97</b>
C.1 Qualtrics print out . . . . .	97
C.2 R markdown for analysis of results . . . . .	115

---

## A Formal Presentation of theory

In what follows, I describe a formal presentation of my trust theory. First, I set up the model, provide definitions of my key terms, and solve for some preliminary results that are useful for the rest of the solution. Second, I consider the case in which the first issue is a peripheral interest. I solve for pure strategy equilibria. Third, I consider the case in which the first issue is a core interest. I solve for pure strategy equilibria. Fourth, I explain how I derive the picture in the manuscript.

### A.1 Set up

I study a model between a Challenger ( $C_j$ , she) and Defender (D, he). The Challenger's type  $j$  represents either greedy  $j = G$  and limited  $j = L$  long-run preferences.

The game tree is presented in Figure 1 of the manuscript. In it the Challenger is given two opportunities for revision  $t \in 1, 2$  unless the Defender intervenes. If the Challenger revises the status quo in either opportunity, she gets  $\pi_t - k_t$  and the Defender gets 0. If she does not revise the status quo, the Challenger gets 0 and the Defender gets 1.

In between these two revision opportunities, the Defender has the option of preventive competition. If the Defender does not prevent, the Challenger is given the second opportunity for revision. If the Defender does prevent, the second opportunity is decided via lottery that  $C$  wins with probability  $p$ . Both players pay a cost  $w$ .

We write the Challengers actions as  $a_t^C \in r_t, nr_t$  if the Challenger selects revision or no revision at  $t$ . We write the Defender's action as  $a^D \in c, nc$  if the Defender selects competition or not.

Utilities are realized at the end of the game. As an example, if the Challenger selects revision, then the Defender competes player utilities are  $U^C(r_1, c) = \pi_1 - k_1 + p\pi_2 - w$ ,  $U^D(r_1, c) = 1 - p - w$ . If instead the Challenger avoids revision in the first period (but not the second) and the Defender does not prevent, then  $U^C(nr_1, r_2, nc) = \pi_2 - k_2$ ,  $U^C(nr_1, r_2, nc) = 1$ .

The information structure of the game follows from three types of random variables that determine the Challenger's costs and benefits from revision. At the beginning of the game, Nature assigns the Challenger as greedy with probability  $pr(j = G) = \alpha$  and limited with probability  $pr(j = L) = 1 - \alpha$ . Then at the beginning of each revision opportunity Nature draws two random variables i.i.d from their respective distributions. First, Nature reveals that the revision opportunity is over the Challenger's declared core interest with probability  $\lambda$  and declared peripheral interest with probability  $1 - \lambda$ . Then Nature draws  $C$ 's cost of revision  $k_t$  from a conditional density function  $f()$  which is supported on the non-negative real numbers. We normalize  $\int_0^\infty f(k)dk = 1$ .

I assume that the Challenger observes every draw as soon as Nature draws it. However, the Defender only observes whether the crisis is over a core or peripheral interest. Therefore, the Defender does not know if the Challenger is limited or greedy, and also does not know the crisis-specific variable  $k_t \sim f()$ .

The draws of  $k_t$  plug directly into the Challenger's utility function. However, the Challenger's value for the issue in dispute depends on the Challenger's type and the revision opportunity. If the Challenger is greedy, the Challenger values all issues  $\pi_t = M$ . If the Challenger holds limited aims, then the Challenger values core interests  $\pi_t = H$  and peripheral interests  $\pi_t = L$ .

**Parameter restrictions:** We make the following parameter restrictions. We make one theoretically motivated assumption:  $H \geq M > L$ . That is, limited aims types care as much or more about their core interests than greedy types do.

We make one assumption for ease:  $L < \max(k_t|f())$  and  $H < \max(k_t|f())$ . Violating these conditions just creates corner conditions in our semi-separating equilibrium. For example, if  $H_t > \max(k_t)$  it means that avoiding revision signals limited aims with certainty under broad conditions.

### A.1.1 Important Definitions

Our focus is on the Defender's beliefs about the Challenger's long-term intentions. Consistent with the existing literature on reassurance, this parameter represents trust. Define  $\alpha_0 = \alpha$  as the Defender's prior belief that the Challenger's aims are limited. D's prior is simply conditional on Nature's draw. Define  $\alpha_1 | \alpha, s^C, s^D, a_1^C, \lambda_1$  as the Defender's belief about the Challenger's motives after the Challenger takes a first period action. The amount that the Defender trusts the Challenger before and after the first period is characterized by  $\alpha_0, \alpha_1$  and the impact of the Challenger's first-period costly military actions on trust is characterized by  $\alpha_1 - \alpha_0$ . Therefore, in any equilibrium in which  $\alpha_1 | r_1 > \alpha_1 | nr_1$  means that a costly military action engenders trust. If instead  $\alpha_1 | r_1 < \alpha_1 | nr_1$  it means that avoiding a costly military action engenders trust. We use the term game-changing trust if the Defender conditions his competition choice on the Challenger's first period intervention choice.

### A.1.2 Preliminaries

Our solution concept is Pure Bayesian Equilibrium (PBE) and we focus on pure strategy equilibria. All the results hinge on what D expects C to do in the second period. So I first analyze the sub-game starting in the second revision opportunity. Suppose, D plays  $nc$ , then C's choice is a simple cost benefit calculus. C fights if  $\pi_2 > k_2$  and does not fight otherwise.

Working backwards, at the point where D must choose to intervene or not, D's expectation that C will select revision at the second opportunity is:

$$pr(a_2^C = r_2) = \alpha_1(\lambda pr(H > k_2) + (1 - \lambda)pr(L > k_2)) + (1 - \alpha_1)pr(M > k_2).$$

It follows that D wants to pay the cost to compete if  $1 - (pr(a_2^C = r_2)) < 1 - p - w \equiv$

$$\alpha_1 < \frac{pr(M > k_2) - p - w}{pr(M > k_2) - pr(L > k_2) - \lambda[pr(H > k_2) - pr(L > k_2)]}$$

and not otherwise.

We can re-write this using  $f()$  as:

$$\alpha_1 < \frac{\int_0^M f(k)dk - p - w}{\int_0^M f()dk - \int_0^L f()dk - \lambda[\int_0^H f()dk - \int_0^L f()dk]}$$

This leads to our trust condition:

$$\alpha_1 < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f()dk - \lambda[\int_L^H f()dk]} \quad (1)$$

This condition defines the amount of trust that D needs to forgo costly competition and give C the second opportunity to enact revision. When the condition is satisfied, D does not trust C enough and selects competition. When it is violated, D does trust C enough and forgoes competition.

Since this condition is very important, we'll take time to explain a few of its features. First, the amount of trust D needs to avoid competition is decreasing in  $\lambda$ . That is, D needs to be quite confident that the limited aims Challenger will face a peripheral interest in the future. Second it is decreasing in L, M and H. In other words, when the chance that we can find a  $k_2$  that exceeds each of these thresholds diminishes, the Defender believes that there are more conditions under which the Challenger will seek second period revision. Finally, the threshold is decreasing in the cost of competition  $w$  and the Challenger's expectation of defeating the Defender if competition comes ( $p$ ).

Our theoretical focus also assumes two facts. First, if the Defender knew that the Challenger held greedy aims that the Defender would want to select competition. This is true if  $1 - \int_0^M f(k)dk < 1 - p - w \implies$

$$\mathcal{A}_1 : \int_0^M f(k)dk > p + w$$

It also implies that if D was certain that C held limited aims that D would prefer to forgo competition and allow the second crisis to continue. This is true iff:  $\lambda(1 - \int_0^H f(k)dk) + (1 - \lambda)(1 - \int_0^L f(k)dk) > 1 - p - w \implies$

$$\mathcal{A}_2 : \lambda < \frac{p + w - \int_0^L f(k)dk}{\int_0^H f(k)dk}$$

Another important part of reassurance is that the Challenger has a preference for avoiding competition. C's expected value for avoiding preventive competition and allowing the second crisis to materialize (i.e. C does not yet know  $\pi_2, k_2$ , but does know her type) depends on her type. If C is the greedy type, then  $\int_0^M f \int_0^M M - kdkdk$ . If C is the Limited type then  $\lambda \int_0^H f \int_0^H H - kdkdk + (1 - \lambda) \int_0^L f \int_0^L L - kdkdk$ .

This leads to our final two assumptions that ensure the Challenger wants to communicate limited aims and avoid the Defender's costly competition:

$$\mathcal{A}_3 := \int_0^M f \int_0^M M - kdkdk > pM - w$$

$$\mathcal{A}_4 := w - pL + \int_0^L f \int_0^L L - kdkdk > \lambda(p(H - L) - (\int_0^H f \int_0^H H - kdkdk - \int_0^L f \int_0^L L - kdkdk))$$

$$\mathcal{A}_4 := \lambda \int_0^H f \int_0^H H - kdkdk + (1 - \lambda) \int_0^L f \int_0^L L - kdkdk > p(\lambda H + (1 - \lambda)L) - w$$

This assumption states that all Challenger types prefer to avoid preventive war and make it to the second crisis. These assumptions expose the reassurance problem. Notice that D will select costly competition if and only if he is confident that the Challenger is the greedy type. If instead, D is confident that the Challenger holds limited aims, D will not compete (i.e. cooperate) because D is sufficiently confident that C does not want to revise the status quo in the second period. However, all Challenger-types want to avoid preventive war. Given D's incentives, the only way that the Challengers do this is by convincing D that they are the limited aims Challenger.

Since Nature draws the features of the first revision opportunity independently from all player actions and the probability of what happens later on, and also that D knows whether the first crisis is over a core or peripheral interest, we can solve for PBE under two conditions that represent two sub-games. The first condition is that the first revision opportunity arises over a peripheral interest. The second condition is that a peripheral interest arises over a core interest.

## A.2 First issue is a peripheral interest

Let's assume that the first revision opportunity is over a declared peripheral interest. We'll now solve for three pure strategy PBE in which the Challenger's choice to avoid revision engenders trust. To be clear, the results in this section add little beyond what we know from the existing literature. Thus, I only briefly discuss them in result 5.

I focus on two kinds of equilibria: Contingent and non-contingent equilibria. These describe whether C's action impacts D's choice to compete.

### A.2.1 Semi-separating Equilibrium with non-contingent threats

We now describe two equilibria where D's choice to compete or not is insensitive to C's action. However, C's decision to avoid revision still engenders a small amount of trust.

**Lemma A.1** *If*

$$\alpha_0 < \frac{\int_M^\infty f(k)dk \times [\int_0^M f(k)dk - p - w]}{\int_L^\infty f(k)dk \times [\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]] - \int_L^M f(k)dk \times [\int_0^M f(k)dk - p - w]} \quad (2)$$

*Then there is a semi-separating pure strategy PBE in which D plays c.  $C_G$  plays  $r_1$  iff  $k_1 < M$ ,  $C_L$  plays  $r_1$  iff  $k_1 < L$ . In this equilibrium C's decision to avoid costly revision always engenders trust:  $\alpha_1|nr_1 > \alpha_1|r_1$ . However, it never engenders enough trust to change D's choice.*

The limited aims Challenger prefers revision if

$$L - k_1 + p(\lambda H + (1 - \lambda)L) - w > 0 + p(\lambda H + (1 - \lambda)L) - w \implies L > k_1$$

The greedy Challenger prefers revision if

$$M - k_1 + pM - w > 0 + pM - w \implies M > k_1$$

These conditions Characterize C's incentives to remain on the path as stated in the equilibrium.

Since  $M > L$  it instantly follows that  $\alpha_1|nr_1 > \alpha_1|r_1$ .

Given their on-path actions,  $\alpha_1|nr_1 = \frac{\alpha_0 \int_L^\infty f(k)dk}{\alpha_0 \int_L^\infty f(k)dk + (1 - \alpha_0) \int_M^\infty f(k)dk}$ .

Also,  $\alpha_1|r_1 = \frac{\alpha_0(1 - \int_L^\infty f(k)dk)}{\alpha_0(1 - \int_L^\infty f(k)dk) + (1 - \alpha_0)(1 - \int_M^\infty f(k)dk)}$ .

Since trust is lower if C selects revision, we can plug  $\alpha_1|nr_1$  into inequality 1. Therefore, Competition is D's best reply:

$$\frac{\alpha_0 \int_L^\infty f(k)dk}{\alpha_0 \int_L^\infty f(k)dk + (1 - \alpha_0) \int_M^\infty f(k)dk} < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]} \quad (3)$$

$$\frac{\alpha_0 \int_L^\infty f(k)dk}{\int_M^\infty f(k)dk + \alpha_0 \int_L^M f(k)dk} < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]} \quad (4)$$

This re-arranges to what is written in the equilibrium. This completes the proof.

We now turn to a second equilibrium that has a similar feel.

**Lemma A.2** *If*

$$\alpha_0 > \frac{[\int_0^M f(k)dk - p - w] \times \int_0^M f(k)dk}{\int_0^L f(k)dk \times [\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]] + \int_L^M f(k)dk \times [\int_0^M f(k)dk - p - w]} \quad (5)$$

*Then there is a semi-separating pure strategy PBE in which D plays nc.  $C_G$  plays  $r_1$  iff  $k_1 < M$ ,  $C_L$  plays  $r_1$  iff  $k_1 < L$ . Given a second opportunity for revision, C plays  $r_2$  iff  $\pi_2 > k_2$ . In this equilibrium C's decision to seek first period costly revision always engenders mistrust:  $\alpha_1|nr_1 > \alpha_1|r_1$ . However, it never engenders enough mistrust to change D's choice.*

The limited aims Challenger prefers revision if

$$L - k_1 + \lambda \int_0^H f \int_0^H H - k dk dk + (1 - \lambda) \int_0^L f \int_0^L L - k dk dk > 0 + \lambda \int_0^H f \int_0^H H - k dk dk + (1 - \lambda) \int_0^L f \int_0^L L - k dk dk$$

$$\implies L > k_1$$

The greedy Challenger prefers revision if

$$M - k_1 + \int_0^M f \int_0^M M - k dk dk > 0 + \int_0^M f \int_0^M M - k dk dk \implies M > k_1$$

Note these are the identical conditions we derived in the other equilibrium in which D plays a non-conditioned strategy. Thus, as with last time, it instantly follows that  $\alpha_1 | nr_1 > \alpha_1 | r_1$  and our expected values for  $\alpha_1 | nr_1, \alpha_1 | r_1$  are the same.

For ease, we can re-write

$$\alpha_1 | r_1 = \frac{\alpha_0 \int_0^L f(k) dk}{\alpha_0 \int_0^L f(k) dk + (1 - \alpha_0) (\int_0^M f(k) dk)}$$

Since trust is higher if C selects no revision, we can plug in  $\alpha_1 | r_1$  into the violated inequality 1. Therefore, no competition is D's best reply if:

$$\frac{\alpha_0 \int_0^L f(k) dk}{\int_0^M f(k) dk - \alpha_0 \int_L^M f(k) dk} > \frac{\int_0^M f(k) dk - p - w}{\int_L^M f() dk - \lambda [\int_L^H f() dk]} \quad (6)$$

This rearranges to the equilibrium condition.

## A.2.2 Semi-Separating Equilibrium with Contingent Threats

We now solve for a pure strategy semi-separating equilibrium where the Defender's competition choice is contingent on the Challenger's first costly revision choice.

In the non-contingent equilibria, the conditions for C to select revision were a simple calculus of  $\pi_1 > k_1$ . In the contingent equilibria, C must factor in the consequences of D's choice. This generates more complicated thresholds for C's first period revision choice. We define them here because they are part of the equilibrium conditions. But we will solve for them in a moment after we state the equilibrium.

Define

$$\underline{k}_1 = \lambda p(H - L) + L(1 + p) - w - \lambda \int_0^H f \int_0^H H - k dk dk - (1 - \lambda) \int_0^L f \int_0^L L - k dk dk \quad (7)$$

$$\bar{k}_1 = M(1 + p) - w - \int_0^M f \int_0^M M - k dk dk \quad (8)$$

We are now ready to characterize the equilibrium.

**Lemma A.3** Assume  $\mathcal{A}_1, \mathcal{A}_2$  and  $0 < \underline{k}_1 < \bar{k}_1 < 1$ . Then if

$$\frac{\alpha_0 \int_0^{\underline{k}_1} f(k) dk}{\int_0^{\underline{k}_1} f(k) dk - \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k) dk} < \frac{\int_0^M f(k) dk - p - w}{\int_L^M f() dk - \lambda [\int_L^H f() dk]} < \frac{\alpha_0 (1 - \int_0^{\underline{k}_1} f(k) dk)}{1 - \int_0^{\underline{k}_1} f(k) dk + \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k) dk} \quad (9)$$

is satisfied, there is a semi-separating pure strategy PBE in which D plays  $c | r_1, nc | nr_1$ .  $C_j$  all play  $r_2$  iff  $\pi_2 > k_2$ .  $C_G$  plays  $r_1$  iff  $k_1 < \bar{k}$ ,  $C_L$  plays  $r_1$  iff  $k_1 < \underline{k}$ . In this separating equilibrium C's decision to avoid costly revision always engenders game-changing trust:  $\alpha_1 | nr_1 > \alpha_1 | r_1$ .

Given D's strategy,  $C_L$  prefers revision to no revision ( $r_1 > nr_1$ ) if:

$$L - k_1 + p(\lambda H + (1 - \lambda)L) - w > 0 + \lambda \int_0^H f \int_0^H H - k dk dk + (1 - \lambda) \int_0^L f \int_0^L L - k$$

This solves for  $k_1 < \lambda p(H - L) + L(1 + p) - w - \lambda \int_0^H f \int_0^H H - k dk dk - (1 - \lambda) \int_0^L f \int_0^L L - k$  and is how we derived  $\underline{k}_1$ .

$C_G$  prefers revision if:

$$M - k_1 + pM - w > 0 + \int_0^M f \int_0^M M - k dk dk$$

This solves for  $k_1 < M(1 + p) - w - \int_0^M f \int_0^M M - k dk dk$  and is how we derived  $\bar{k}_1$ .

We can use these  $k$  cut-points to solve for  $\alpha_1$ .

Specifically, if C revises, then

$$\alpha_1|r_1 = \frac{\alpha_0 \int_0^{\underline{k}_1} f(k) dk}{\alpha_0 \int_0^{\underline{k}_1} f(k) dk + (1 - \alpha_0) \int_0^{\bar{k}_1} f(k) dk}$$

$$\alpha_1|r_1 = \frac{\alpha_0 \int_0^{\underline{k}_1} f(k) dk}{\int_0^{\bar{k}_1} f(k) dk - \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k) dk}$$

If C does not then

$$\alpha_1|nr_1 = \frac{\alpha_0 \int_{\underline{k}_1}^1 f(k) dk}{\alpha_0 \int_{\underline{k}_1}^1 f(k) dk + (1 - \alpha_0) \int_{\bar{k}_1}^1 f(k) dk}$$

$$\alpha_1|nr_1 = \frac{\alpha_0 \int_{\underline{k}_1}^1 f(k) dk}{\alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k) dk + \int_{\bar{k}_1}^1 f(k) dk}$$

Given that  $C_G$ 's threshold never exceeds  $C_L$ 's it must be that  $\alpha_1|r_1 < \alpha_1|nr_1$  as stated in the equilibrium.

Using the trust condition 1 D prefers to make competition contingent on C's first period revision choice rather than deviate if:

$$\alpha_1|r_1 < \frac{\int_0^M f(k) dk - p - w}{\int_L^M f(k) dk - \lambda[\int_L^H f(k) dk]} < \alpha_1|nr_1$$

Plugging in our results gets is condition 9 as stated in the equilibrium.

### A.3 Violent revision engenders trust when faced with a core interest crisis

We now complete the same exercise under the assumption that the first revision opportunity arises for a core interest. These are the results mainly discussed in the manuscript.

### A.3.1 Semi-separating Equilibrium with non-contingent threats

We now describe two equilibria where D's choice to compete or not is insensitive to C's action. Unlike the peripheral interest case, and in contrast to the existing literature, we'll show that C's revision engenders a small amount of trust.

**Lemma A.4** *If*

$$\alpha_0 < \frac{\int_M^\infty f(k)dk \times [\int_0^M f(k)dk - p - w]}{\int_H^\infty f(k)dk \times [\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]] + \int_M^H f(k)dk \times [\int_0^M f(k)dk - p - w]} \quad (10)$$

*Then there is a semi-separating pure strategy PBE in which D plays c. C<sub>G</sub> plays r<sub>1</sub> iff k<sub>1</sub> < M, C<sub>L</sub> plays r<sub>1</sub> iff k<sub>1</sub> < H. In this equilibrium C's decision to engage in costly revision always engenders trust: α<sub>1</sub>|nr<sub>1</sub> > α<sub>1</sub>|r<sub>1</sub> if H > M. However, costly revision cannot engenders enough trust to change D's competition choice.*

The limited aims Challenger prefers revision if

$$H - k_1 + p(\lambda H + (1 - \lambda)L) - w > 0 + p(\lambda H + (1 - \lambda)L) - w \implies H > k_1$$

The greedy Challenger prefers revision if

$$M - k_1 + pM - w > 0 + pM - w \implies M > k_1$$

These conditions Characterize C's incentives to remain on the path as stated in the equilibrium.

Since  $H \geq M$  it instantly follows that  $\alpha_1|r_1 \geq \alpha_1|nr_1$ .

Given their on-path actions,  $\alpha_1|nr_1 = \frac{\alpha_0 \int_H^\infty f(k)dk}{\alpha_0 \int_H^\infty f(k)dk + (1 - \alpha_0) \int_M^\infty f(k)dk}$ .

Also,  $\alpha_1|r_1 = \frac{\alpha_0 \int_0^H f(k)dk}{\alpha_0 (\int_0^H f(k)dk) + (1 - \alpha_0) \int_0^M f(k)dk}$ .

Since trust is lower if C selects revision, we can plug  $\alpha_1|nr_1$  into inequality 1. Therefore, competition is D's best reply if:

$$\frac{\alpha_0 \int_H^\infty f(k)dk}{\alpha_0 \int_H^\infty f(k)dk + (1 - \alpha_0) \int_M^\infty f(k)dk} < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]} \quad (11)$$

$$\frac{\alpha_0 \int_H^\infty f(k)dk}{\int_M^\infty f(k)dk - \alpha_0 \int_M^H f(k)dk} < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]} \quad (12)$$

This re-arranges to what is written in the equilibrium. This completes the proof.

We now turn to a second equilibrium that has a similar feel.

**Lemma A.5** *If*

$$\alpha_0 > \frac{[\int_0^M f(k)dk - p - w] \times \int_0^M f(k)dk}{\int_0^L f(k)dk \times [\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]] - \int_L^M f(k)dk \times [\int_0^M f(k)dk - p - w]} \quad (13)$$

*Then there is a semi-separating pure strategy PBE in which D plays nc. C<sub>G</sub> plays r<sub>1</sub> iff k<sub>1</sub> < M, C<sub>L</sub> plays r<sub>1</sub> iff k<sub>1</sub> < H. In this equilibrium C's decision to avoid costly revision always mistrust: α<sub>1</sub>|nr<sub>1</sub> > α<sub>1</sub>|r<sub>1</sub> if H > M. However, avoiding revision never engenders enough trust to change D's competition choice.*

As we've shown previously, The limited aims Challenger prefers revision if  $H > k_1$  and the greedy Challenger prefers revision if  $M > k_1$ . As with last time, it instantly follows that  $\alpha_1|nr_1 > \alpha_1|r_1$  and our expected values for  $\alpha_1|nr_1, \alpha_1|r_1$  are the same.

Since trust is higher if C selects no revision, we can plug in  $\alpha_1|r_1$  into the violated inequality 1. Therefore, competition is D's best reply if:

$$\frac{\alpha_0 \int_0^H f(k)dk}{\int_0^M f(k)dk + \alpha_0 \int_M^H f(k)dk} > \frac{\int_0^M f(k)dk - p - w}{\int_L^M f()dk - \lambda[\int_L^H f()dk]} \quad (14)$$

This rearranges to the equilibrium condition.

### A.3.2 Semi-Separating Equilibrium with Contingent Threats

We now solve for a pure strategy semi-separating equilibrium where the Defender's competition choice is contingent on the Challenger's first costly revision choice. But unlike the peripheral interest crisis, we'll show that the Challenger engenders trust by fighting and therefore the Defender responds with competition if and only if the challenger avoids a costly revisionist attempt over her core interest.

As in the peripheral interest case, in the contingent equilibria C must factor in the consequences of D's choice into her decision to signal trust through revision. This generates more complicated thresholds for C's first period revision choice. We define them here because they are part of the equilibrium conditions. But we will solve for them in a moment after we state the equilibrium.

Define

$$\underline{k}_1 = M(1 - p) + w + \int_0^M f \int_0^M M - kdkdk \quad (15)$$

$$\bar{k}_1 = H(1 - p\lambda) - (1 - \lambda)L + w + \lambda \int_0^H f \int_0^H H - kdkdk + (1 - \lambda) \int_0^L f \int_0^L L - kdkdk \quad (16)$$

These conditions serve the same function as they did in the contingent equilibrium when the first revision opportunity arose over a peripheral interest. However, in this case  $\underline{k}_1$  is associated with  $C_g$ 's threshold and  $\bar{k}_1$  is associated with  $C_L$ ' threshold. This difference emphasizes that the contingent equilibrium requires that the limited aims type will revise the status quo over a core interest with higher values of  $k_1$ .

We are now ready to characterize the equilibrium.

**Lemma A.6** *If*

$$0 < \underline{k}_1 < \bar{k}_1 < 1 \quad (17)$$

*and*

$$\frac{\alpha_0 \int_{\underline{k}_1}^{\infty} f(k)dk}{\int_{\underline{k}_1}^{\infty} f(k)dk - \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k)dk} < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f()dk - \lambda[\int_L^H f()dk]} < \frac{\alpha_0 \int_0^{\bar{k}_1} f(k)dk}{\int_0^{\bar{k}_1} f(k)dk + \alpha_0 \int_{\bar{k}_1}^{\infty} f(k)dk} \quad (18)$$

*is satisfied. Then, there is a semi-separating pure strategy PBE in which D plays  $c|nr_1, c|r_1$ .  $C_j$  all play  $r_2$  iff  $\pi_2 > k_2$ .  $C_G$  plays  $r_1$  iff  $k_1 < \underline{k}$ ,  $C_L$  plays  $r_1$  iff  $k_1 < \bar{k}$ . In this separating equilibrium C's decision to engage in costly revision always engenders game-changing trust:  $\alpha_1|r_1 > \alpha_1|nr_1$ .*

Given D's strategy,  $C_L$  prefers revision to no revision ( $r_1 > nr_1$ ) if:

$$H - k_1 - w + \lambda \int_0^H f \int_0^H H - kdkdk + (1 - \lambda) \int_0^L f \int_0^L L - kdkdk > 0 + p(\lambda H + (1 - \lambda)L)$$

$$k_1 < H(1 - p\lambda) - (1 - \lambda)L + w + \lambda \int_0^H f \int_0^H H - kdkdk + (1 - \lambda) \int_0^L f \int_0^L L - kdkdk$$

This is how we derived  $\bar{k}_1$   
 $C_G$  prefers revision if:

$$M - k_1 + \int_0^M f \int_0^M M - k dk dk > 0 + pM - w$$

$$k_1 < M(1 - p) + w + \int_0^M f \int_0^M M - k dk dk$$

This is is how we derived  $\underline{k}_1$ .

As stated in equilibrium condition 17, the equilibrium only holds together if  $C_L$ 's threshold exceeds  $C_G$ 's threshold. This is true if:

$$H(1 - p\lambda) - (1 - \lambda)L + w + \lambda \int_0^H f \int_0^H H - k dk dk + (1 - \lambda) \int_0^L f \int_0^L L - k dk dk > M(1 - p) + w + \int_0^M f \int_0^M M - k dk dk$$

$$H(1 - p\lambda) - L(1 - \lambda) - M(1 - p) + \lambda \int_0^H f \int_0^H H - k dk dk + (1 - \lambda) \int_0^L f \int_0^L L - k dk dk - \int_0^M f \int_0^M M - k dk dk > 0$$

Unfortunately, I could not confirm that this was always the case. However, we will show later that it can be satisfied under the assumption that there is deep uncertainty over the Challenger's issue-specific value (e.g  $f = U[0, 1]$ ). Using partial derivatives we can also confirm that the left hand side is strictly increasing in  $H$  and strictly decreasing in  $M$  for any  $f(\cdot)$ . Furthermore, a simple substitution confirms that we cannot satisfy this if  $\lambda < 1/2$  and  $M = H$ .

This leads to the following remark that bounds my theoretical claim:

**Remark** The contingent threats semi-separating equilibrium in which violent revision engenders game changing trust requires that the limited aims Challenger value core interests more than the greedy Challenger values all issues. As  $H - M$  increases, it is easier to find conditions under which different kinds of Challengers prefer their on-path play.

To be clear, this is not the case for the semi-separating equilibrium with non-contingent threats we discussed in the last sub-section. If  $\alpha_0$  is sufficiently low (high) such that no matter what C does at the first revision opportunity D will always select (avoid) competition, then revision always engenders trust.

With this caveat aside, we proceed to derive the remaining equilibrium conditions under the assumption that 17 holds. If it does hold, then we can be certain that fighting engenders some amount of trust. We can use the  $k$  thresholds to derive exactly how much. Specifically, if C revises, then

$$\alpha_1 | r_1 = \frac{\alpha_0 \int_0^{\bar{k}_1} f(k) dk}{\alpha_0 \int_0^{\bar{k}_1} f(k) dk + (1 - \alpha_0) \int_0^{\underline{k}_1} f(k) dk}$$

$$\alpha_1 | r_1 = \frac{\alpha_0 \int_0^{\bar{k}_1} f(k) dk}{\int_0^{\underline{k}_1} f(k) dk + \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k) dk}$$

If C does not then

$$\alpha_1 | nr_1 = \frac{\alpha_0 \int_{\bar{k}_1}^{\infty} f(k) dk}{\alpha_0 \int_{\bar{k}_1}^{\infty} f(k) dk + (1 - \alpha_0) \int_{\underline{k}_1}^{\infty} f(k) dk}$$

$$\alpha_1 | nr_1 = \frac{\alpha_0 \int_{\bar{k}_1}^{\infty} f(k) dk}{\int_{\underline{k}_1}^{\infty} f(k) dk - \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k) dk}$$

Using the trust condition 1 D prefers to make competition contingent on C's first period revision choice rather than deviate if:

$$\alpha_1|nr_1 < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f()dk - \lambda[\int_L^H f()dk]} < \alpha_1|r_1$$

Plugging in our results,

$$\frac{\alpha_0 \int_{\underline{k}_1}^{\infty} f(k)dk}{\int_{\underline{k}_1}^{\infty} f(k)dk - \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k)dk} < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f()dk - \lambda[\int_L^H f()dk]} < \frac{\alpha_0 \int_0^{\bar{k}_1} f(k)dk}{\int_0^{\bar{k}_1} f(k)dk + \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k)dk} \quad (19)$$

#### A.4 Generating the Comparative static plot in the Manuscript

We generate plots using the following assumptions:  $H = .9p = .5$ ,  $w = .05$ ,  $L = .1\lambda = .15\alpha_0 = .3$

**Preliminaries** We showed above that D's best response (to take up or avoid competition) depends on how much D trusts C. We derived a trust condition that we re-write given our assumptions about  $f()$  as:

$$\alpha_1 < \frac{M - \theta(p - w)}{M - L - \lambda(H - L)}$$

This states that if D's posterior belief that C holds limited aims ( $\alpha_1$ ) is low enough, then D prefers competition.

We now solve the specific equilibria using this trust condition.

##### A.4.1 C' revision signals non-game changing trust. D always competes

**Summary of Lemma:** Lemma A.4: D competes,  $C_G$  plays  $r_1$  iff  $k_1 < M$ ,  $C_L$  plays  $r_1$  iff  $k_1 < H$ .

**Thresholds for C's first period revision choice:** In this case, the Challenger takes a revisionist action in the first period if  $k_1 < \pi_1$ . This means that the greedy Challenger takes on revision if  $k_1 < M$  and the limited aims Challenger takes on revision if  $k_1 < H$ . The reason that this choice is so simple is that D does not condition his action on C's choice. Therefore, C makes a simple cost-benefit calculus knowing that it will not impact the future.

**Posterior Beliefs given on path action:**  $\alpha_1|nr_1 = \frac{\alpha_0(1-H/\theta)}{\alpha_0(1-H/\theta) + (1-\alpha_0)(1-M/\theta)}$ .

$$\alpha_1|nr_1 = \frac{\alpha_0(1 - H/\theta)}{1 - M/\theta - \alpha_0(H - M)/\theta}$$

$$\alpha_1|r_1 = \frac{\alpha_0 H}{\alpha_0 H + (1-\alpha_0)M}$$

**The condition for the equilibrium to hold**  $\alpha_0 < \theta \frac{(1-M/\theta) \times [M/\theta - p - w]}{(1-H/\theta) \times [(M-L) - \lambda[H-L]] + (H-M) \times [M/\theta - p - w]}$

This states that D starts out with prior beliefs that C is very likely the greedy Challenger. Thus, C's choice does not impact D's beliefs enough to change D's behavior. Of course, this does not mean that D does not learn from C's actions. It simply means that learning does not impact D's strategy.

##### A.4.2 C's revision signals non-game changing trust. D never competes.

**Summary of Lemma:** Lemma A.5: D no compete,  $C_G$  plays  $r_1$  iff  $k_1 < M$ ,  $C_L$  plays  $r_1$  iff  $k_1 < H$ .

**Thresholds for C's first period revision choice:** In this case, the Challenger takes a revisionist action in the first period if  $k_1 < \pi_1$ . This means that the greedy Challenger takes on revision if  $k_1 < M$  and the limited aims Challenger takes on revision if  $k_1 < H$ . The reason that this choice is so simple is that D does

not condition his action on C's choice. Therefore, C makes a simple cost-benefit calculus knowing that it will not impact the future.

**Posterior Beliefs Given C's on path action**  $\alpha_1|nr_1 = \frac{\alpha_0(1-H/\theta)}{\alpha_0(1-H/\theta)+(1-\alpha_0)(1-M/\theta)}$ .

$\alpha_1|r_1 = \frac{\alpha_0 H}{\alpha_0 H + (1-\alpha_0)M}$ .

**The Condition for the equilibrium to hold.**

$$\alpha_0 > \frac{[M/\theta - p - w] \times M}{L \times [M - L - \lambda[H - L]]/\theta - (M - L) \times [M/\theta - p - w]} \quad (20)$$

This states that D starts out with prior beliefs that C is very likely a limited aims Challenger. Thus, C's choice does not impact D's beliefs enough to change D's behavior. Of course, this does not mean that D does not learn from C's actions. It simply means that learning does not impact D's strategy.

### A.5 C's decision to fight engenders game changing trust:

**Summary of Lemma:** Lemma A.6: D  $c|nr_1, nc|r_1$ .  $C_L$  plays  $r_1$  iff  $k_1 < \bar{k}_1$ ,  $C_G$  plays  $r_1$  iff  $k_1 < \underline{k}_1$ .

**Thresholds:** The lower threshold belongs to the greedy type. The Greedy type takes the first revision opportunity if:

$$\underline{k}_1 < M(1 - p) + w + M^2/2\theta \quad (21)$$

$$\underline{k}_1 < M(1 - p + M/2\theta) + w \quad (22)$$

The upper threshold belongs to the limited aims type.  $C_L$  takes the first revision opportunity if

$$\bar{k}_1 < H(1 - p\lambda) - (1 - \lambda)L + w + \lambda H^2/2\theta + (1 - \lambda)L^2/2\theta \quad (23)$$

$$\bar{k}_1 < H(1 - p\lambda + \lambda H/2\theta) - L(1 - \lambda)(1 - L/2\theta) + w \quad (24)$$

**Threshold condition:** The equilibrium requires that  $\bar{k}_1 > \underline{k}_1$ :

$$H(1 - p\lambda + \lambda H/2\theta) > L(1 - \lambda)(1 - L/2\theta) + M(1 - p + M/2\theta)$$

**Posterior Beliefs:** If D observes C select revision,

$$\alpha_1|r_1 = \frac{\alpha_0 \times [H(1 - p\lambda + \lambda H/2\theta) - L(1 - \lambda)(1 - L/2\theta) + w]}{M(1 - p + M/2\theta) + w + \alpha_0 \times [H(1 - p\lambda + \lambda H/2\theta) - L(1 - \lambda)(1 - L/2\theta) + w]}$$

If D observes no revision:

$$\alpha_1|nr_1 = \frac{\alpha_0[1 - H(1 - p\lambda + \lambda H/2\theta) + L(1 - \lambda)(1 - L/2\theta) - w]}{1 - M(1 - p + M/2\theta) - w - \alpha_0[H(1 - p\lambda + \lambda H/2\theta) - L(1 - \lambda)(1 - L/2\theta) - M(1 - p + M/2\theta)]}$$

**Posterior beliefs incentives D to condition competition:** The following conditions parameterize the general statement in condition 25.

The first condition for the equilibrium to hold is that if D observes no revision, D is so confident that C is greedy that D turns to competition:

$$\frac{\alpha_0[1 - H(1 - p\lambda + \lambda H/2\theta) + L(1 - \lambda)(1 - L/2\theta) - w]}{1 - M(1 - p + M/2\theta) - w - \alpha_0[H(1 - p\lambda + \lambda H/2\theta) - L(1 - \lambda)(1 - L/2\theta) - M(1 - p + M/2\theta)]} < \frac{M - \theta(p - w)}{M - L - \lambda(H - L)}$$

The second condition

$$\frac{M - \theta(p - w)}{M - L - \lambda(H - L)} < \frac{\alpha_0 \times [H(1 - p\lambda + \lambda H/2\theta) - L(1 - \lambda)(1 - L/2\theta) + w]}{M(1 - p + M/2\theta) + w + \alpha_0 \times [H(1 - p\lambda + \lambda H/2\theta) - L(1 - \lambda)(1 - L/2\theta) + w]}$$

$$\frac{\alpha_0 \int_{\bar{k}_1}^{\infty} f(k)dk}{\int_{\underline{k}_1}^{\infty} f(k)dk - \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k)dk} < \frac{\int_0^M f(k)dk - p - w}{\int_L^M f(k)dk - \lambda[\int_L^H f(k)dk]} < \frac{\alpha_0 \int_0^{\bar{k}_1} f(k)dk}{\int_0^{\underline{k}_1} f(k)dk + \alpha_0 \int_{\underline{k}_1}^{\bar{k}_1} f(k)dk} \quad (25)$$

## A.6 Exploring the Boundaries of violent trust.

In this section we ask: how can we adjust the model so that violent trust is not possible? Answering this question connects our model to the classic result.

### A.6.1 Identifying why my result is different from the conventional wisdom: Limited aims Challengers value core interests as much or more than greedy Challengers ( $H \geq M$ ).

The principal modeling innovation was that limited aims Challenges could care as much  $M = H$  or more  $H > M$  about core interests than greedy Challengers. We identified the conditions under which violence engenders trust under that assumption.

The standard assumption is that limited aims Challengers always value issues less in expectation. Given that we include uncertainty over  $k_1, k_2$ , then the closest presentation to the standard model is  $M > H = L$ . However, if we accept that different states can value territories for different reasons, then it is plausible that  $M > H \geq L$  represents the standard model.

We'll now show that violent trust relies on breaking this assumption. This result is important because it connects our results to the standard model in which violent revision always engenders mistrust.

**Proposition A.7** *Assume that  $H < M$  but all other features of the model are the same and assumptions  $\mathcal{A}_1 - \mathcal{A}_4$  hold. Then, there is no equilibrium where C's decision to fight engenders trust. However, there are conditions under which PBE exist where violence engenders mistrust.*

There are two cases to consider: the contingent and non-contingent case. In the non-contingent case, D's strategy does not depend on C's strategy. Thus, C fights in the first period if  $v(\pi_1) - k_1 + EU_2 > EU_2$ . Here  $EU_2$  represents C's expected value in the second period. Since D's choice is insensitive to C's first period action, then  $EU_2$  is identical in both conditions. It follows that C fights in the first period if  $v(\pi_1) > k_1$  and not otherwise.

Clearly, if  $H < M$  then there are more realized values of  $k_1$  for which the greedy type fights. This implies that violent revision must engender mistrust.

Turning to the contingent case. In this case, D selects preventive competition with larger probability if C avoids first-period competition. There must be a realized value of  $k_1$  for which the Greedy Challenger prefers to fight but the limited type does not.

Arguing by contradiction, consider the limiting case  $M = H$  and suppose this  $k_1$  exists. Then for a fixed  $k_1$  the greedy type prefers to backdown rather than fight in the first crisis given that D will avoid competition:

$$M - k_1 + (M - e(k_2)|k_2 > M) < Mp - w$$

$$M - Mp + w + (M - e(k_2)|k_2 > M) < k_1$$

For the same  $k_1$  it must also be the case that the limited aims Challenger prefers to fight and avoid preventive competition:

$$M - k_1 + \lambda(M - e(k_2)|k_2 > M) + (1 - \lambda)(L - e(k_2)|k_2 > L) > \lambda Mp + (1 - \lambda)Lp - w$$

$$M + w - \lambda Mp - (1 - \lambda)Lp + \lambda(M - e(k_2)|k_2 > M) + (1 - \lambda)(L - e(k_2)|k_2 > L) > k_1$$

Putting them together,

$$M + w - \lambda Mp - (1 - \lambda)Lp + \lambda(M - e(k_2)|k_2 > M) + (1 - \lambda)(L - e(k_2)|k_2 > L) > M - Mp + w + (M - e(k_2)|k_2 > M)$$

Which means:

$$\lambda(M - e(k_2)|k_2 > M) + (1 - \lambda)(L - e(k_2)|k_2 > L) > (M - e(k_2)|k_2 > M) - MP$$

This cannot be true if  $\lambda < 1$ , a contradiction.

## A.7 Defenders are uncertain if a territory is a genuine core/peripheral interests.

I assumed that Defenders know which territories fit the Challenger's core interests and which fit peripheral interests with certainty.

My historical review of cases shows that Defenders frequently used these methods to 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<sup>40</sup> (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 world. 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.<sup>41</sup> Consistent 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 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

<sup>40</sup>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.

<sup>41</sup>Author's interview with Former Deputy Secretary Amb. Robert Kimmit; who led the assessment.

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.

If D's intelligence about C's core interests is imperfect, can we still support violent trust? To address this question, I adjust the model as follows. During the first revision opportunity, C privately observes if the issue is a core interest. D correctly observes if the issue is core/peripheral with probability  $\psi \in [.5, 1]$  and incorrectly observes whether the issue is a core or peripheral interest with probability  $1 - \psi$ . If  $\psi$  is high, it means that D employs cultural experts that well understand that a specific issue fits and do not fit C's core interests. If  $\psi = .5$ , it means that D has no expertise and therefore is totally unsure if the issue in dispute is a core or peripheral interest.

The result follows closely from the main model. In every case, if  $\psi$  is sufficiently close to 0.5 then I cannot support violent trust. But as  $\psi \rightarrow 1$  then the results converge to the violent trust equilibria described in the main model. To illustrate the intuition, I provide one example using a non-contingent equilibrium. But the result obviously extends to the contingent equilibrium for sufficiently large  $H$ ,  $\psi$ .

**Lemma A.8** *Consider the model where D is uncertain if the first issue fits C's core interests. Assume that D observes the first issue over a core interest, then the strategies described in the violent trust equilibrium from lemma A.5 is an equilibrium if instead of condition 13, the following is satisfied:*

$$\frac{\alpha_0 \int_0^M f(k) dk}{\alpha_0 \int_0^M f(k) dk + (1 - \alpha_0)[(1 - \psi)(1 - \lambda) \int_0^L f(k) dk + \psi \lambda \int_0^H f(k) dk]} < \frac{\int_0^M f(k) dk - p - w}{\int_L^M f(k) dk - \lambda[\int_L^H f(k) dk]} \quad (26)$$

Since D does not condition his strategy, C's incentives are identical to the proofs for the lemmas referenced from the baseline model. The only question is if D's beliefs support a non-conditional response.

D's posterior belief that C is greedy is:

$$\alpha_1 | r_1, \text{ core} = \frac{\alpha_0 pr(M > k_1)}{\alpha_0 pr(M > k_1) + (1 - \alpha_0)[(1 - \psi)(1 - \lambda) pr(L > k_1) + \psi \lambda pr(H > k_1)]} \quad (27)$$

Notice that this differs from the baseline model in the numerator. Like the baseline model, D's posterior belief weights the possibility that C holds limited aims and fought for a core interest. But unlike the baseline model, it also includes a term that represents the possibility that C holds limited aims, but the issue in dispute was a peripheral interest and D fought for it anyway.

Plugging this value into D's trust threshold, we get condition 26.

**Remark** Condition 26 converges to condition 13 as  $\psi \rightarrow 1$  from below.

Notice that that D's posterior belief defined in 27 is identical to the main model if  $\psi = 1$ . The reason is that the term  $(1 - \psi)(1 - \lambda) pr(L > k_1)$  drops out. It follows that  $\psi < 1$  necessarily degrades the amount that D learns from C's choice to pursue violent revision.

**Remark** Condition 26 ensures that some information is necessary to support violent trust ( $\psi > .5$ ).

To see this is the case, I set  $\psi = 1/2$  in 26. Arguing by contradiction, suppose that we could support violent trust in this condition, then  $\alpha_1 < \alpha_0 \implies \frac{\alpha_0 pr(M > k_1)}{\alpha_0 pr(M > k_1) + \frac{(1 - \alpha_0)^2}{2} [(1 - \lambda) pr(L > k_1) + \lambda pr(H > k_1)]} < \alpha_0$ .

This solves for  $1 < \alpha_0$ , a contradiction.

## B Main (Japan) experiment

First, I provide an anonymous copy of the materials contained in my pre-registration. This includes a pre-analysis plan, the Qualtrics print-out of the survey instrument, and pre-registered r code.

Second, I include an R markdown file that runs the analysis and generates the figures in the manuscript. It also includes summary statistics and other information.

Before we begin, I highlight how my study complies with APSA's principals of engaging human subjects in research. This information applies to both studies.

In general, I comply with and support all of the principals described by APSA. Both studies were part of one IRB submission to my institutional review board. That submission was deemed exempt for review by my institutional IRB. There is no deception in either study.

As I describe in the PAP and as I present in the qualtrics print-out, subjects were provided with consent form that they could agree to (leading to the survey) or not (leading to termination). This consent form provides the details of research, my contact information, the contact information of my IRB, etc.

I recruited subjects via Lucid. Lucid is a well recognized survey firm and subjects self-enrol on its Theorem platform then select tasks to complete. I paid Lucid \$1 and they have a standard (undisclosed) rate that they pay recruited participants. Subjects click on a link that re-directs them to a survey (with an online, IRB-approved consent form) that I programmed in Qualtrics. Subjects then complete my survey. Qualtrics estimated the average survey time at 9.5 minutes. See the PAP for more details.

To account for recent inattentiveness in online responses, I used two pre-treatment attention checks. Consistent with Lucid's rules, the IRB forms, and APSA ethical standards, I dropped and did not pay subjects from the sample who fail either of these checks. These checks ask subjects about information presented in the prior slide. One question asked subjects to pick the two names of Islands that were mentioned on the last slide (it lists pairs of Islands which which subjects chose). The other asks subjects to identify the potential interests of Japan listed in the last slide.<sup>42</sup> The exact wording of the questions is listed in the instrument below.

The data taken from the elite interviews was collected as part of a related but different project while I was employed at a different institution. The elites I interviewed were conducted before this project was conceived. Nevertheless, I sought and obtained IRB approval as part of that initial project and the information I used fit within the guidelines of the information management plan registered at that IRB. The interviews complied with APSA Principles. Subjects were emailed an IRB approved consent form before the interview. When they asked, I mailed them a copy of the interview template before the interview. Subjects were told that they would be recorded, and that quotes from the the records would be used in academic research publications. Subjects were told that the recordings would not be released without their expressed consent. Subjects were asked if I could attribute quotes to them and I only attribute quotes with their consent. Subjects are high-level government officials with top secret security clearances and decades of public interview experience. I did not compensate them. There was no deception.

### B.1 Pre-analysis Plan

---

<sup>42</sup>The supplemental study used similar questions but replaced the Japan-specific information with the generic information used in that study.

# Pre-Analysis Plan

March 30, 2022

# 1 Introduction

In what follows I summarize the technical information that relates to my implementation procedure.

A detailed explanation for my theory and how I derive my hypotheses, including a formal model, is attached as the Theory Appendix.

The qualtrics print-out for my survey instrument appears as the Qualtrics Appendix.

The Rcode for my data manipulations, pre-registered test for my main hypothesis, and my proposed visualizations of my results appear in the attached R markdown file.

In what follows I describe other features of my analysis plan including sampling procedures, solicitation and also the inferences I draw based on specific statistical tests.

## 2 Implementation procedure

I will generate evidence to test my theory use a survey experiment embedded in a foreign policy scenario.

I plan to recruit subjects via lucid online survey platform and program the instrument in Qualtrics. Lucid sets payment at \$1. Qualtrics estimates the average survey time at 9.5 minutes. To account for recent inattentiveness in online responses, I used two pre-treatment attention checks. Consistent with Lucid's rules, I can drop these subjects from the sample. I also time the survey. I define my fully attentive sample as those who pass these two attention checks and take longer than 3.5 minutes (210 sec) to complete the survey. My plan is to analyze the fully attentive sample.<sup>1</sup>

I recruit 200 subjects and assign them i.i.d to treatment arms with a random variable ratio of 2:1:1 respectively.<sup>2</sup>

Lucid provides demographic information. But I will not use it to draw inferences.

I plan to analyze the data using R statistical software. I compute permutation tests using the coin package. I generate plots using the ggplot2 package. My code for downloading the data, manipulating it, and my statistical tests and some plots are described in an R markdown document appended to this analysis plan.

Below I provide a discussion of the salient features for drawing inferences.

---

<sup>1</sup>As I explain below, I also include a post-treatment attention check. But my goal is not to use it.

<sup>2</sup>I duplicate the first treatment arm in Qualtrics.

## 2.1 Using Japan to code variation in historical and instrumental value

My experiments focus on a fictional Japan in 2040. Japan has unique historical features that allowed me to precisely construct scenarios that measures my core variables and tease out the implications of my theory. Japan's current sovereign borders includes 5 main Islands and many smaller Islands. There are also many other Islands that Japan has some interest in, but that Japan does not control. These smaller Islands through the East Asian Seas allow me to isolate territories that are important to Japan for different reasons. I exploit variation in this context to test different parts of my theory.

I code the Amami Islands as a core interest that scores high in historical and cultural importance to Japan but low in instrumental value.

I code the Senkaku Islands as a core interest that scores moderate in normative and strategic value.

I code the Liancourt Islands as a peripheral interest that scores low in normative importance and moderate in strategic value.

## 2.2 Vignette

The vignette has three phases. In phase 1, all subjects are provided factual information about Japan's foreign policy history, territorial borders, and Japan's connection to different Islands.

In phase 2, all subjects are provided plausible information about Japan's foreign policy trajectory over the next two decades to set up a classic trust problem. Notably, Japan loses the Amami and Senkaku Islands to Korea. Also, Japan is emerging in a period of military and economic strength at a time its neighbors are weak.

Phase 2 provides plausible assessments of Japan's potential foreign policy motives. The final page of the prompt summarizes two potential sets of motives that Japan could hold. These match both my theory and the history of Japan's foreign policy preferences under different regimes. It states,

Japan's foreign policy statement explains that Japan has limited territorial ambitions: "During the 2030s, several Islands of vital historical and cultural significance were stolen from us. We will re-take Islands of historical significance through force if necessary. However, we hold no other territorial ambitions."

A Chinese diplomat disagrees: "Japan's appeal to limited aims is a cover for its true ambition: regional domination. At first, Japan will selectively expand into territories of economic and

strategic value to consolidate power. If Japan succeeds at this first step, we won't be strong enough to stop Japan from dominating all of East Asia."

At the end of phase 2, I measure subjects prior (i.e. pre military action) level of trust using a metric I describe in a moment.

In phase 3, subjects are randomly assigned into one of three crises over different Islands:

- Treatment A: Japan invades the Amami Islands.
- Treatment B: Japan invades the Senkaku Islands
- Treatment C: Japan invades the Liancourt Island

At the end of phase 3, I measure subjects posterior (i.e. post randomly assigned military action) level of trust using a metric I describe in a moment.

### **2.3 Measuring Trust.**

I measure trust using the following question

We now want your opinion about how likely these outcomes are relative to each other. In summary, here are the two possibilities:

A. Limited: Japan will at most expand its borders to take Islands of historical and cultural importance.

B. Expansive: Japan will opportunistically expand its borders across East Asia. Japan may not always fight. But it will expand when the strategic and economic value outweighs the cost.

How likely are these outcomes relative to each other?

Subjects are presented with a slide rule that goes from -10 to 10. The slide rule is labeled at -10 = Definitely Limited; 10 = Definitely Expansive and 0 = equally likely. The slide rule starts at 0 and responses are coerced to integer values.

When exported to R, responses are recorded as integers  $\in -10 : 10$ .

## 2.4 Pre-registered tests

My main prediction is that violent revision over a core interest engenders trust. This is a relative claim. And we can state it relative to two counter-factuals.

The first counter-factual is relative to time.

**H1:** Subjects that observe Japan use violent military force to re-take the Amami Islands will increase their trust that Japan's long run intentions are limited.

I test this hypothesis by analyzing how the crisis treatment impacts subject trust scores who are assigned to the Amami Island scenario. I subset the results to treatment arm A and measure the difference in means for prior and posterior trust.

I use the p-value derived from a one-sided permutation test to infer support for my hypothesis (see R code). If the pre-treatment mean  $<$  post-treatment mean of and the p-value derived from the permutation test confirms that the means are distinguishable with 90% confidence, I infer moderate support for H1. If the p-value derived from the permutation test confirms that the means are distinguishable with 95% confidence, I infer support for H1. If the p-value derived from the permutation test confirms that the means are distinguishable with 99% confidence, I infer strong support for H1.

See R code for more details.

**H2:** Subjects that observe Japan use violent military force to re-take the Amami Islands will trust that Japan's long run intentions are limited more than subjects who observe Japan use violent military force to take the Liancourt Islands.

I use the p-value derived from a one-sided permutation test to infer support for my hypothesis (see R code). If the mean of treatment group A  $<$  mean of treatment group C, and the p-value derived from the permutation test confirms that the means are distinguishable with 90% confidence, I infer moderate support for H2. If the p-value derived from the permutation test confirms that the means are distinguishable with 95% confidence, I infer support for H2. If the p-value derived from the permutation test confirms that the means are distinguishable with 99% confidence, I infer strong support for H2.

### 2.4.1 Visualizing data

Although I pre-register permutation tests, I plan to report visualized data in the manuscript (see R code for the most likely plot). That plot will be a 2x3 plot. Where the first dimension represents prior v. posterior

trust measure and the second dimension is treatment arm. So long as the visualized results accurately reflect the statistical tests, I will report the sample means and the 95% confidence intervals.

## 2.5 Other evidence

I now describe other evidence I collect and initial plans to analyze it.

### 2.5.1 Alternative to H2

A different way to adjudicate H2 is to show that different groups update in different directions. In large samples, the prior beliefs should be invariant across treatment arms. But in small samples, it is possible to find that the baseline scores vary across treatment groups because a handful of outliers. Factoring in the baseline estimates could help account for this random variability.

To compute this test, I first compute the difference in the posterior and prior trust estimates. I then use a permutation test to estimate whether the means of this computed value are different across treatment groups A and C. I use the same 90, 95, 99 p-values described above to infer moderate, normal and strong levels of support.

### 2.5.2 A notable conjecture

I plan to examine one conjecture.

**Conjecture:** Observing Japan use violent military force to re-take the Senkaku Islands will not affect a subject's level of trust over time.

I write this as a conjecture rather than an explicit hypothesis because because I could not precisely measure subject perceptions about how sensitive different Challenger-types are to different sources of value.

In fact, if I find evidence that fighting over the Senkaku Islands generates some trust, then it would illustrate the far-reaching implications of my theory because it would suggest that even when territories have huge economic value, that subjects interpret the normative value as the dominant component for all kinds of states.

As a result, I treat this conjecture as exploring the bounds of the domain of violence and reassurance.

### 2.5.3 Text responses

At the end of phase three, we ask subjects whether the Amami Islands Crisis impacted their assessment of Japan.

We intend to categorize these results as consistent with violent trust, consistent with conventional wisdom, consistent with core interest wars have no affect, other.

### 2.5.4 Post-survey questions

I record three types of post-survey questions that help validate my measures and explore our causal mechanism.

These questions are exploratory. They are designed to help tease out the logic behind the results that I observe. Thus, I cannot derive firm predictions about how I will analyze them because I don't know what the experimental results will be.

**Validating my trust measure** My outcome measure emphasizes two likely territorial ambitions Japan could hold. There could be others. To check whether mine are the most likely I ask:

In the scenario above we presented two possibilities about Japan's territorial ambition:

A. Limited: Japan will at most expand its borders into territories of historical and cultural importance. It will then stop taking territory.

B. Expansive: Japan will opportunistically expand its borders across East Asia. Japan may not fight Japan will expand when the strategic and economic value outweighs the cost. .

We now want to know if that characterization was plausible. Please select the option that best fits your opinion.

- The presented outcomes were the only plausible outcomes
- The presented outcomes were more plausible than other outcomes
- There was a third outcome that was just as likely.
- There was a third outcome that was more likely than the two presented.

This question helps me establish that my primary trust question captures the two main possible ambitions of Japan.

I say the outcomes I focus on are highly salient if 75% of subjects answer one of the first two answers. I say the outcomes I focus on are the most salient if 60% of subjects answer with one of the first two answers. I say the outcomes I focus on are salient if 50% of subjects answer with one of the first two answers.

To be clear, how subjects answer this question will not bias my answers. In fact, the very reason I focused on two outcomes in my trust question (limited =-5, expansive =5) ensures that alternatives I fail to consider do not bias my inferences. If I focus on a single outcome, (e.g do you think that Japan's aims are limited) I introduce potential bias.

**Game-Changing Trust** Result 4 of my model suggests that trust can impact competition choices. But the result depends heavily on prior beliefs  $\alpha_0$ , as well as perceptions of the costs of conflict  $f \sim k$ . Thus, I don't want to make strong predictions.

But for interest, I ask subjects about whether the US should sell weapons to Japan.

Very strong support for my theory would suggest that subjects who observed Japan fight over the Amami Islands would be more supportive of this sale than subjects who observed Japan fight over the Liancourt Islands.

**Self-reported core interests** I ask subjects if they believe the issue in dispute served Japan's core interests. In general, I expect that the distribution of these responses should match my self-coded values.

If it doesn't it could help explain my inability to identify affects.

Further, I expect self-reporting on this question should be correlated with trust.

### 2.5.5 Post-survey attentive checks

After the treatment, I ask a third attention check I ask subjects what Japan did. They can answer Japan could invade / not invade. Clearly, those who select not invade did not gather the relevant information.

If over 10% of the sample fail this question, this could explain substantial noise in the results. As a result, I reserve the right to sample additionally, or include this check in the attentive sample should I get high failure rates on this question.

But I would reduce my confidence to moderate if I did.

## References

## **B.2 Qualtrics print out**

## **consent**

### **Consent Form.**

This is a study meant to gauge public opinion on issues of U.S. foreign policy including military conflict. My name is [REDACTED] and I am a researcher at [REDACTED]. Participation in this study will involve completing the following survey. The survey will take approximately 10 minutes.

There are no known or anticipated risks to you for participating in the survey. The benefit to you comes in the form of the compensation you were offered when you were recruited into the survey. Only respondents who correctly answer attention check questions will be allowed to complete the survey. Failure to pass this check may affect compensation. In addition, we hope that this survey can also lead to a better understanding about important issues U.S. politicians must consider. All of your responses will be anonymous.

Only the researchers involved in this study and those responsible for research oversight will have access to the

information you provide. Your responses will be recorded in a dataset. No identifying information will be collected or stored in our dataset.

Participation in this study is completely voluntary. You are free to decline to participate, to end participation at any time for any reason, or to refuse to answer any individual question. Your decision whether or not to participate in this study will not affect your relationship with [REDACTED], or LUCID.

If you have any questions about this study, you may contact the researchers [REDACTED] [REDACTED]. If you would like to talk with someone other than the researcher to discuss problems or concerns, to discuss situations in the event that the researcher is not available, or to discuss your rights as a research participant, you may contact the [REDACTED] [REDACTED] additional information is available at [REDACTED]

Are you willing to participate in this study?

- Yes
- No

## **prompt**

The following pages describe a plausible future foreign policy scenario involving Japan. Similar hypothetical scenarios are used to form US foreign policy. Please take it seriously.

We present:

1. Factual information about Japan's history.
2. Plausible information about Japan's trajectory in 2040.

We then ask you to assess Japan's interest in territorial conquest in 2040.

## **factual**

### **Factual Information About Japan.**

Japan consists of 5 main Islands and 100s of smaller Islands. Japan values these smaller Islands for different

reasons. For example:

### The Senkaku Islands:

- Claimed by Japan in 1890.
- Uninhabited.
- Huge oil deposits.

### The Amami Islands:

- Settled by Japan in 625
- Home to 200,000 Japanese citizens / important cultural sites.
- No natural resources.

We just mentioned two smaller Islands that Japan controls.

They were:

- Kumi and Senkaku
- Senkaku and Amami
- Amami and Okinawa
- Okinawa and Kumi

**prior**

## **How East Asia unfolds: 2022-2039.**

Imagine the year is 2040.

- Between 2028-2035 Japan fought a civil war.
- Japan's neighbors seized Japan's smaller Islands. Korea invaded the Amami and Senkaku Islands and took them from Japan.
- In 2036, the United States and China fought a devastating war. The United States withdrew forces from East Asia.
- In 2036, a charismatic leader reunited Japan's five main Islands under a new, single-party Government.
- Japan's economy started booming. It is now the largest economy in Asia.

Now in 2040, Japan is entering a period of military strength at a time its neighbors are weaker. Analysts agree: Japan

could take territory across East Asia if it wanted to. The question is: What territory does Japan want to take?

Japan's government explains that its has limited territorial ambitions:

*“During the 2030s, several Islands of vital historical significance were stolen from us. We will re-take Islands of historical significance, through force if necessary. However, we hold no other territorial ambitions.”*

A Chinese diplomat disagrees:

*“Japan’s appeal to limited aims is a cover for its true ambition: regional domination. At first, Japan will selectively expand into territories of economic and strategic value to consolidate power. If Japan succeeds at this first step, we won't be strong enough to stop Japan from dominating all of East Asia.”*

We just described two kinds of foreign ambition that Japan could hold. Pick the two that best fit what we described.

Japan will not seek any territorial revision

- Japan's territorial ambition is limited to Islands of historical importance
- Japan seeks regional domination in East Asia.

## prior\_response

Correct!

We now want your opinion about how likely these outcomes are relative to each other. In summary, here are the two possibilities:

A. **Limited**: Japan will at most expand its borders to take Islands of historical value.

B. **Expansive**: Japan will opportunistically expand its borders across East Asia. Japan may not always fight. But it will expand when the strategic and economic value outweighs the cost.

How likely are these two outcomes relative to each other?



Definitely Limited

Equally Likely

Definitely  
Expansive

-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10



## Amami\_crisis

### A crisis over the Amami Islands.

Korea is constructing military bases on the Amami Islands. Korea plans to build 10,000 homes on these Islands and populate them with Korean nationals.

As stated earlier, Korea invaded the Amami Islands in 2035 and took them from Japan.

For Japan, the Amami Islands hold:

- Huge historical value
- Little economic/strategic value.

### Japan's 2 options:

The Japanese leadership considered the following report from the Japanese military:

1. If we do not invade the Amami Islands, Korea will populate them with Korean nationals, and consolidate power.
2. If we invade the Amami Islands, we will certainly prevail. However, invasion will be costly. We will suffer casualties, and lose ships and aircraft that we would need if we wanted to complete other military missions.

**Decision:**

Japan invaded the Amami Islands.

Japan re-took these Islands of vital historical importance but little strategic importance. Japan suffered casualties and lost material that it would use for other military missions.

Japan's Foreign Minister claimed that "Our actions confirm that our territorial ambition is limited to Islands of historical importance."

## **Amami\_2**

### **A crisis over the Amami Islands.**

Korea is constructing military bases on the Amami Islands. Korea plans to build 10,000 homes on these Islands and populate them with Korean nationals.

As stated earlier, Korea invaded the Amami Islands in 2035 and took them from Japan.

For Japan, the Amami Islands hold:

- Huge historical value
- Little economic/strategic value.

### **Japan's 2 options:**

The Japanese leadership considered the following report from the Japanese military:

1. If we do not invade the Amami Islands, Korea will populate them with Korean nationals, and consolidate power.

2. If we invade the Amami Islands, we will certainly prevail. However, invasion will be costly. We will suffer casualties, and lose ships and aircraft that we would need if we wanted to complete other military missions.

## **Decision:**

Japan invaded the Amami Islands.

Japan re-took these Islands of vital historical importance but little strategic importance. Japan suffered casualties and lost material that it would use for other military missions.

Japan's Foreign Minister claimed that "Our actions confirm that our territorial ambition is limited to Islands of historical importance."

## **Senkaku\_crisis**

### **A crisis over the Senkaku Islands.**

Korea is constructing military bases on the Senkaku Islands. Korea plans to build 10,000 homes on these Islands and populate them with Korean nationals.

As stated earlier, in 2035 Korea invaded the Senkaku Islands and took them from Japan.

For Japan, the Senkaku Islands have:

- historical value
- economic/strategic value.

### **Japan's 2 options:**

The Japanese leadership considered the following report from the Japanese military:

1. If we do not invade the Senkaku Islands, Korea will populate them, and consolidate power. We may never get them back.

2. If we invade the Senkaku Islands, we will certainly prevail. However, invasion will be costly. We will suffer casualties, and lose ships and aircraft that we would need if we wanted to complete other military missions.

### **Decision:**

Japan invaded the Senkaku Islands.

Japan re-took these Islands of historical and economic value. Japan suffered casualties and lost material that it would use for other military missions.

Japan's Foreign Minister claimed that "Our actions confirm that our territorial ambition is limited to Islands of historical importance."

### **Liancourt\_crisis**

**A crisis over the Liancourt Island.**

Korea is constructing military bases on the Liancourt Island. Korea plans to build 10,000 homes on this Island and populate it with Korean nationals.

Korea has usually controlled the Liancourt Island. However, Japan briefly controlled it in the 16th century.

For Japan, the Liancourt Island has:

- A small amount of historical value
- A moderate amount of economic/strategic value.

## **Japan's 2 options:**

The Japanese leadership considered the following report from the Japanese military:

1. If we do not invade the Liancourt Island, Korea will populate it, and consolidate power. We may never get it back.
2. If we invade the Liancourt Island, we will certainly prevail. However, invasion will be costly. We will suffer casualties,

and lose ships and aircraft that we would need if we wanted to complete other military missions.

## **Decision:**

Japan invaded the Liancourt Island.

Japan took this Island that holds a little historical value and some strategic value. Japan suffered casualties and lost material that it would use for other military missions.

Japan's Foreign Minister claimed that "Our actions confirm that our territorial ambition is limited to Islands of historical importance."

## **post\_response**

With this new information, we want you to re-assess the relative likelihood of the following two possibilities:

A. **Limited:** Japan will at most expand its borders to take Islands of historical value.

B. **Expansive**: Japan will opportunistically expand its borders across East Asia. Japan may not always fight. But it will expand when the strategic and economic value outweighs the cost.

How likely are these two outcomes relative to each other?



Did the crisis event influence your assessment? If it did, why?

Which option best summarizes Japan's crisis response:

- Invaded Island
- Did not invade Island

The crisis you observed was over an Island. Which description best describes Japan's claims over the Island

- The Island in dispute clearly fell within Japan's core interests.
- The Island in dispute likely fell within Japan's core interests
- It was difficult to tell if the Island in dispute served Japan's core interests
- The Island in dispute was clearly not one of Japan's core interests

## ps\_2

### **Were these two outcomes plausible?**

In the scenario above we presented two possibilities about Japan's territorial ambition:

A. *Limited*: Japan will at most expand its borders into territories of historical importance. It will then stop taking territory.

B. *Expansive*: Japan will opportunistically expand its borders across East Asia. Japan may not fight Japan will

expand when the strategic and economic value outweighs the cost.

We now want to know if that characterization was plausible. Please select the option that best fits your opinion.

- The presented outcomes were the only plausible outcomes
- The presented outcomes were more plausible than other outcomes
- There was a third outcome that was just as likely.
- There was a third outcome that was more likely than the two presented.

If there was a plausible third outcome, what was it?

**ps\_1**

## **US Weapons sales?**

Japan's Navy wants to purchase 100 combat ships from United States. These ships are effective for naval patrols. They could also be used in operations to defend or take Islands. They are ineffective for taking territory in mainland Asia.

Weapon sales are important for the US economy and could also engender trust with Japan.

Should the United States sell Japan these vessels?

- Definitely not
- Probably not
- Might or might not
- Probably yes
- Definitely yes

### **B.3 Pre-registered analytical procedure in R**

# R code for main experiment

## Violent Trust code to analyze for experiment

The following is the pre-registered code for the main experiment as part of the project on violent trust.

```
rm(list = ls())
require(tidyverse)
require(qualtRics)
library("gridExtra")
library(coin)
library(plyr)

#####
##### Download data from Qualtrics
#####

readRenviron("~/Renviron")

# Fetching specific surveyresults.
response <- fetch_survey(surveyID = XXXX,
  verbose = TRUE, convert = TRUE, label = TRUE, force_request = T)

#####
# Cleaning data
#####

glimpse(response)

response$treatment <- NA
response$treatment[which(response$FL_32_DO_Amami_crisis==1)] <- "Amami"
response$treatment[which(response$FL_32_DO_Amami_2==1)] <- "Amami"
# As stated in PAP, I over-assign subjects into Amami group.
response$treatment[which(response$FL_32_DO_Senkaku_crisis==1)] <- "Senkaku"
response$treatment[which(response$FL_32_DO_Liancourt_crisis ==1)] <- "Liancourt"

response$difference <- response$post_response_1 - response$prior_response_1

table(response$treatment)

complete <- response %>% filter(response$treatment >0) %>%
  filter(`Duration (in seconds)` > 210)
# This is my "fully attentive sample" that I intend to analyze.
```

```

#####
# Analysis: Hypothesis 1
#####

complete %>% filter(treatment=="Amami") %>%
  # Focus only on Amami sub-group.
  select(treatment,post_response_1,prior_response_1) %>%
  pivot_longer(cols = post_response_1:prior_response_1,
               names_to = "timing",
               values_to = "value") %>%
  mutate(timing = as.factor(timing)) %>%
  independence_test(value ~ timing, data = .,
ytrafo = rank_trafo, distribution = exact(), alternative="less")

# p-value < .05 I infer support
# p-value < .01 I infer strong support
# p-value < .1 I infer moderate support
# for my theory.

#####
# Analysis: Hypothesis 2
#####

complete %>% filter(treatment=="Amami"|treatment=="Liancourt") %>%
  # Focus only on Amami /Liancourt
  mutate(treatment = as.factor(treatment)) %>%
  independence_test(post_response_1 ~ treatment,
data = ., ytrafo = rank_trafo, distribution = exact(),alternative="less")

# p-value < .05 I infer support
# p-value < .01 I infer strong support
# p-value < .1 I infer moderate support
# for my theory.

#####
# Analysis: Conjecture 1
#####

##Since I predict a Null, I test each of the one-sided tests separately. My conjecture is that there is

complete %>% filter(treatment=="Senkaku") %>% # Focus only on Senkaku sub-group.
  select(treatment,post_response_1,prior_response_1) %>%
  pivot_longer(cols = post_response_1:prior_response_1,
               names_to = "timing",
               values_to = "value") %>%
  mutate(timing = as.factor(timing)) %>%
  independence_test(value ~ timing, data = .,
ytrafo = rank_trafo, distribution = exact())

complete %>% filter(treatment=="Senkaku") %>% # Focus only on Amami sub-group.

```

```

select(treatment,post_response_1,prior_response_1) %>%
pivot_longer(cols = post_response_1:prior_response_1,
              names_to = "timing",
              values_to = "value") %>%
mutate(timing = as.factor(timing)) %>%
independence_test(value ~ timing, data = .,
ytrafo = rank_trafo, distribution = exact(), alternative="less")

#####
# Likely visualization of data.

### Write functions#####
summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE,
                      conf.interval=.80, .drop=TRUE) {
  library(plyr)

  # New version of length which can handle NA's: if na.rm==T, don't count them
  length2 <- function(x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else      length(x)
  }

  # This does the summary. For each group's data frame, return a vector with
  # N, mean, and sd
  datac <- ddply(data, groupvars, .drop=.drop,
                .fun = function(xx, col) {
                  c(N      = length2(xx[[col]], na.rm=na.rm),
                    mean   = mean  (xx[[col]], na.rm=na.rm),
                    sd     = sd    (xx[[col]], na.rm=na.rm)
                  )
                },
                measurevar
  )
  datac$se <- datac$sd / sqrt(datac$N) # Calculate standard error
  # Confidence interval multiplier for standard error
  # Calculate t-statistic for confidence interval:
  # e.g., if conf.interval is .95, use .975 (above/below), and use df=N-1
  ciMult <- qt(conf.interval/2 + .5, datac$N-1)
  datac$ci <- datac$se * ciMult
  datac$name <- as.factor(measurevar)
  return(datac)
}

# First, I create a summary data table broken out by treatment group and pre/post treatment
su.data <- rbind(summarySE(complete,
measurevar="prior_response_1", groupvars=c("treatment"), na.rm = TRUE),
                 summarySE(complete,
measurevar="post_response_1", groupvars=c("treatment"), na.rm = TRUE))

```

```

su.data$low.ci <- su.data$mean - su.data$ci
su.data$high.ci <- su.data$mean + su.data$ci

#####
# This plot reports the mean and 95% Confidence intervals of prior and posterior.
#####
ggplot(su.data, aes(x=mean, y=treatment, color = name) ) +
  geom_point(size = 4 , position = position_dodge(width = 0.3)) +
  geom_errorbarh(aes(xmin = low.ci ,
xmax = high.ci, height = 0), size=1.1, alpha=.5,
position = position_dodge(width = 0.3)) +
  theme_classic() +
  theme(legend.position = "bottom") + labs(x = NULL, title = NULL, y=NULL)

#####
# Analysis: Checking post-survey question
#####
### Did subjects beleive that the two presented outcomes were the most likely?
table(complete$plausible_outcomes )

#####
# Post-treatment attention check
#####

## I also included a post-treatment attention check since the instrument is complex.
# This just asks subjects if they know Japan invaded, which is constant across all treatment groups.
#If subjects cannot answer this question correctly, then it shows they lost attention
# before they rendered their main answer. Therefore, their answer only adds noise.

complete2 <- response %>% filter((Attention_3 == "Invaded Island" )) %>%
  filter(`Duration (in seconds)` > 210)

# If 10% of subjects fail this attention check, and failure is similar across groups,
# This would point to lost attention as one reason the survey failed.
# This may cause me to analyze this sub-sample.
# See PAP for details.

#####
# Alternative H2.
#####
# Sometimes with small samples, it is possible that the group priors are substantially different.
#This could make it hard to recover a differenced affects.
### An alternative way to test this hypotheses is to measure the difference (between treatment groups)
#

```

```

complete %>% filter(treatment=="Amami"|treatment=="Liancourt") %>% # Focus only on Amami /Liancourt
  mutate(treatment = as.factor(treatment)) %>%
  independence_test(difference ~ treatment, data = ., ytrafo = rank_trafo, distribution = exact(), alterna

# p-value < .05 I infer support
# p-value < .01 I infer strong support
# p-value < .1 I infer moderate support
# for my theory.

# This is the associated plot:

# First, I create a summary data table broken out by treatment group and pre/post treatment
su.data2 <- rbind(summarySE(complete,
measurevar="difference", groupvars=c("treatment"), na.rm = TRUE))

su.data2$low.ci <- su.data2$mean - su.data2$ci
su.data2$high.ci <- su.data2$mean + su.data2$ci

# Second, I plot the mean and 95% Confidence intervals using the sample standard error.
ggplot(su.data2, aes(x=mean, y=treatment) ) +
  geom_point(size = 4 , position = position_dodge(width = 0.3)) +
  geom_errorbarh(aes(xmin = low.ci , xmax = high.ci,
height = 0), size=1.1, alpha=.5, position = position_dodge(width = 0.3)) +
  theme_classic() +
  theme(legend.position = "bottom") + labs(x = NULL, title = NULL, y=NULL)

```

## **B.4 R markdown for analysis of results**

# Running Pre-registered Code and Generating plots for paper

In this document I take the pre-registered code for the main experiment and run it with the completed survey results. I print out the data tables I used to generate the figures and then run the code I used to generate the figures. I also run a number of alternative permutation tests mentioned in the pre-registered documentation but that were not pre-registered as main tests.

## Loading Packages and downloading data:

```
rm(list = ls())
require(tidyverse)

## Loading required package: tidyverse
## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.3      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
require(qualtRics)

## Loading required package: qualtRics
## Warning: package 'qualtRics' was built under R version 4.1.2
library(coin)

## Warning: package 'coin' was built under R version 4.1.2
## Loading required package: survival
library(plyr)

## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
```

```

## The following object is masked from 'package:purrr':
##
## compact
library(ggribbles)

readRenviron("~/Renviron")
surveys <- all_surveys()
response <- fetch_survey(surveyID = surveys$id[10],
verbose = TRUE, convert = TRUE, label = TRUE, force_request = T)

## |
##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   StartDate = col_datetime(format = ""),
##   EndDate = col_datetime(format = ""),
##   Status = col_character(),
##   IPAddress = col_character(),
##   Finished = col_logical(),
##   RecordedDate = col_datetime(format = ""),
##   ResponseId = col_character(),
##   RecipientLastName = col_logical(),
##   RecipientFirstName = col_logical(),
##   RecipientEmail = col_logical(),
##   ExternalReference = col_logical(),
##   DistributionChannel = col_character(),
##   UserLanguage = col_character(),
##   consent_yn = col_character(),
##   attention_1 = col_character(),
##   attention_2_1 = col_character(),
##   attention_2_2 = col_character(),
##   attention_2_3 = col_character(),
##   post_text = col_character(),
##   Attention_3 = col_character()
##   # ... with 14 more columns
## )
## i Use `spec()` for the full column specifications.
## Warning in file.remove(survey.fpath): cannot remove file 'C:/Users/micha/
## AppData/Local/Temp/Rtmp4gtsVC/Japan_3_islands_200_publish.csv', reason
## 'Permission denied'

```

## Cleaning data

The first thing I need to do is define a `treatment` variable. As mentioned in pre-registration, I double sampled on the Amami-Island treatment (which is high historical) to increase my confidence I could validate H1. I also generate a variable `difference` that measures changing estimates over time.

```

response$treatment <- NA
response$treatment[which(response$FL_32_DO_Amami_crisis==1)] <- "Amami"
response$treatment[which(response$FL_32_DO_Amami_2==1)] <- "Amami"
response$treatment[which(response$FL_32_DO_Senkaku_crisis==1)] <- "Senkaku"

```

```
response$treatment[which(response$FL_32_DO_Liancourt_crisis ==1)] <- "Liancourt"
response$difference <- response$post_response_1 - response$prior_response_1
```

Focusing on the attentive sample:

```
complete <- response %>% filter(is.na(response$treatment)==F) %>%
  filter(`Duration (in seconds)` > 210)
table(complete$treatment)
```

```
##
##      Amami Liancourt   Senkaku
##      97      41      47
```

The table shows that randomization worked well!

## Analysis

We will now test our pre-registered hypotheses.

### Analysis: Hypothesis 1

H1 asked if subjects in the High historical group (Amami) updated over time. We now run the pre-registered test. Pre-registration defined three potential levels of support: support ( $p < .05$ ), strong support ( $p < .01$ ), moderate support ( $p < 0.1$ ).

```
complete %>% filter(treatment=="Amami") %>% # Focus only on Amami sub-group.
  select(treatment,post_response_1,prior_response_1) %>%
  pivot_longer(cols = post_response_1:prior_response_1,
               names_to = "timing",
               values_to = "value") %>%
  mutate(timing = as.factor(timing)) %>%
  independence_test(value ~ timing,
data = ., ytrafo = rank_trafo, distribution = exact(), alternative="less")
```

```
##
## Exact General Independence Test
##
## data: value by
## timing (post_response_1, prior_response_1)
## Z = -1.997, p-value = 0.02287
## alternative hypothesis: less
```

We infer support from our hypothesis.

### Hypothesis 2

H2 asked if subjects in the High historical group (Amami) had different posterior beliefs than the strategic group (Liancourt). We now run the pre-registered test. Pre-registration defined three potential levels of support: support ( $p < .05$ ), strong support ( $p < .01$ ), moderate support ( $p < 0.1$ ).

```
complete %>% filter(treatment=="Amami"|treatment=="Liancourt") %>%
  mutate(treatment = as.factor(treatment)) %>%
  independence_test(post_response_1 ~ treatment,
data = ., ytrafo = rank_trafo, distribution = exact(), alternative="less")
```

```
##
## Exact General Independence Test
##
## data: post_response_1 by treatment (Amami, Liancourt)
## Z = -2.2756, p-value = 0.01126
## alternative hypothesis: less
```

We infer support

## Conjecture

I pre-registered a conjecture about one treatment we believed scored moderate on historical and strategic dimensions (Senkaku). I predict that in ambiguous settings that no updating happens. If violence engendered mistrust in ambiguous settings, then the existing literature on violence and mistrust seems to cover most of the real world cases.

Since I draw different inferences for positive and negative updating, I pre-registered one-sided permutation tests separately.

```
complete %>% filter(treatment=="Senkaku") %>% # Focus only on Amami sub-group.
  select(treatment,post_response_1,prior_response_1) %>%
  pivot_longer(cols = post_response_1:prior_response_1,
               names_to = "timing",
               values_to = "value") %>%
  mutate(timing = as.factor(timing)) %>%
  independence_test(value ~ timing, data = ., ytrafo = rank_trafo, distribution = exact())
```

```
##
## Exact General Independence Test
##
## data: value by
## timing (post_response_1, prior_response_1)
## Z = -0.82995, p-value = 0.4094
## alternative hypothesis: two.sided
```

```
complete %>% filter(treatment=="Senkaku") %>% # Focus only on Amami sub-group.
  select(treatment,post_response_1,prior_response_1) %>%
  pivot_longer(cols = post_response_1:prior_response_1,
               names_to = "timing",
               values_to = "value") %>%
  mutate(timing = as.factor(timing)) %>%
  independence_test(value ~ timing, data = ., ytrafo = rank_trafo, distribution = exact(), alternative=
```

```
##
## Exact General Independence Test
##
## data: value by
## timing (post_response_1, prior_response_1)
## Z = -0.82995, p-value = 0.2047
## alternative hypothesis: less
```

I confirm the conjecture. I do not find a statistically significant difference. However, the results clearly show that I am closer to finding an affect for violent trust than violent mistrust in this case.

## Visualization of data.

I now describe how I visualize the data. As described in my pre-registration documents, if I pass my pre-registered tests, I would use the sample means and 95% confidence intervals in my plots.

The function I wrote to compute the means and sample standard deviations (which is based on a common function on stack exchange) is as follows.

```
summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE,
                      conf.interval=.80, .drop=TRUE) {
  library(plyr)

  # New version of length which can handle NA's: if na.rm==T, don't count them
  length2 <- function(x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else      length(x)
  }

  # This does the summary. For each group's data frame, return a vector with
  # N, mean, and sd
  datac <- ddply(data, groupvars, .drop=.drop,
                .fun = function(xx, col) {
                  c(N    = length2(xx[[col]], na.rm=na.rm),
                    mean = mean  (xx[[col]], na.rm=na.rm),
                    sd   = sd    (xx[[col]], na.rm=na.rm)
                  )
                },
                measurevar
  )
  datac$se <- datac$sd / sqrt(datac$N) # Calculate standard error
  ciMult <- qt(conf.interval/2 + .5, datac$N-1)
  datac$ci <- datac$se * ciMult
  datac$name <- as.factor(measurevar)
  return(datac)
}
```

The data table to generate the ladder plot comes from this code:

```
su.data <- summarySE(complete, measurevar="prior_response_1", na.rm = TRUE)
su.data$treatment <- "Prior"
su.data$id <- NULL
su.data <- rbind(su.data,
                summarySE(complete, measurevar="post_response_1", groupvars=c("treatment"), na.rm = TRUE))

su.data$low.ci <- su.data$mean - su.data$ci
su.data$high.ci <- su.data$mean + su.data$ci
su.data$treatment <- as.factor(su.data$treatment)
su.data$treatment <- factor(su.data$treatment, levels = c("Prior", "Amami", "Liancourt", "Senkaku"))

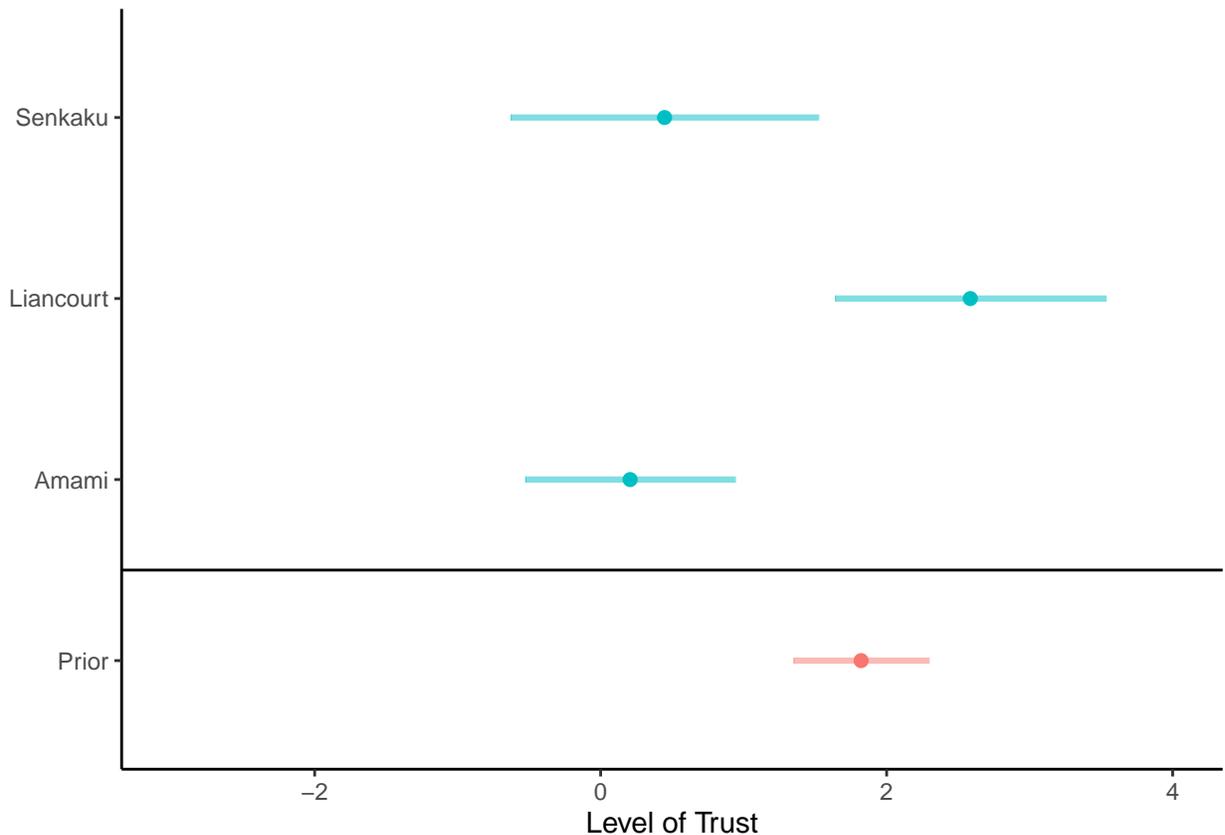
su.data
```

```
##      N      mean      sd      se      ci      name treatment
## 1 185 1.8216216 4.997344 0.3674120 0.4725541 prior_response_1  Prior
## 2  97 0.2061856 5.597503 0.5683404 0.7334048 post_response_1  Amami
## 3  41 2.5853659 4.647449 0.7258096 0.9457858 post_response_1 Liancourt
## 4  47 0.4468085 5.671471 0.8272691 1.0756385 post_response_1  Senkaku
##      low.ci  high.ci
```

```
## 1  1.3490675 2.2941757
## 2 -0.5272193 0.9395904
## 3  1.6395801 3.5311516
## 4 -0.6288300 1.5224470
```

I then made the ladder plot using ggplot.

```
ggplot(su.data, aes(x=mean, y=treatment, color = name) ) +
  geom_point(size = 2 , position = position_dodge(width = 0.3)) +
  geom_errorbarh(aes(xmin = low.ci , xmax = high.ci, height = 0), size=1.1, alpha=.5, position = position_dodge(width = 0.3)) +
  theme_classic() +
  theme(legend.position = "NULL") + labs(x = "Level of Trust", title = NULL, y=NULL) +
  xlim(-3, 4) +
  geom_hline(yintercept = 1.5, linetype="solid", color = "black", size=.5)
```



```
#ggsave("ladder.pdf", width = 8, height = 4)
```

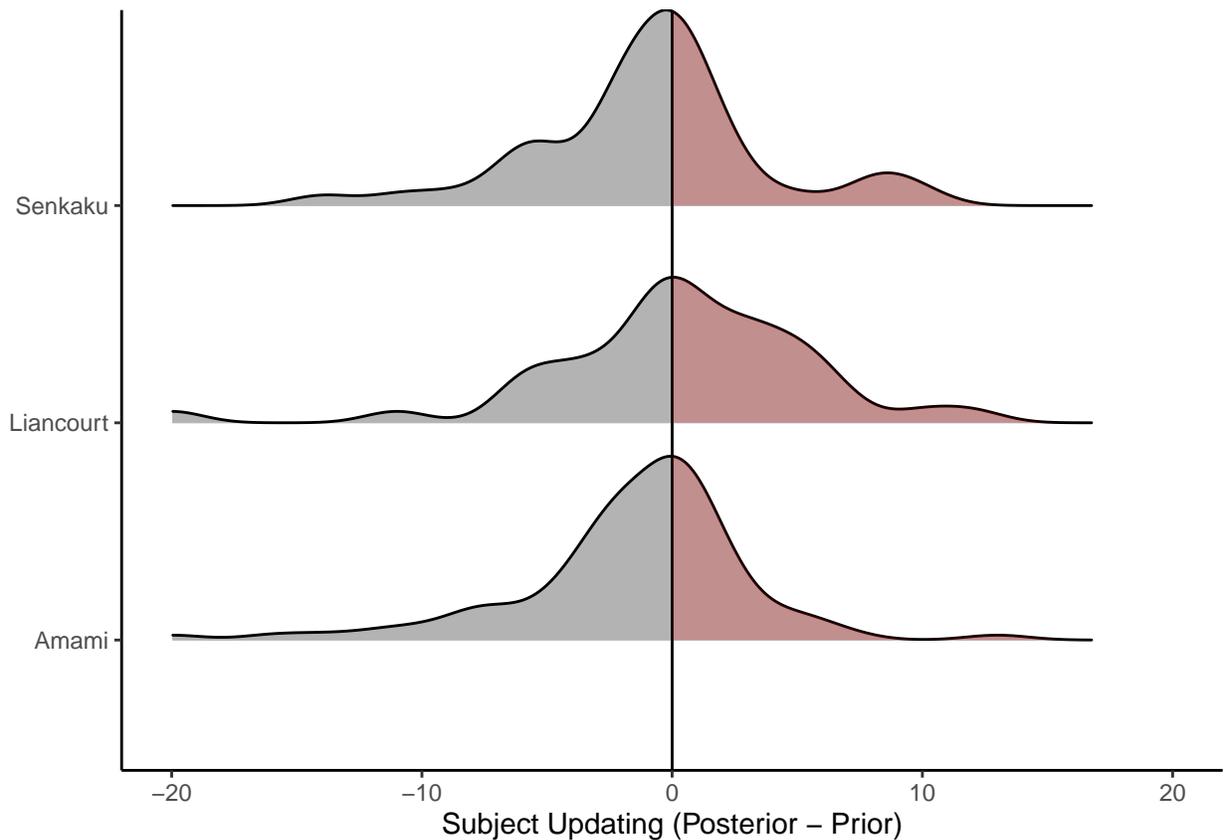
Finally, I generated the density plots as follows

```
gg <- ggplot(complete, aes(x = difference, y = treatment)) +
  stat_density_ridges(
    geom = "density_ridges_gradient", scale = .9) +
  theme_classic() +
  geom_vline(xintercept = 0, linetype="solid", color = "black", size=.5)

# Build ggplot and extract data
d <- ggplot_build(gg)$data[[1]]
```

```
## Picking joint bandwidth of 1.25
# Add geom_ribbon for shaded area
gg +
  geom_ribbon(
    data = transform(subset(d, x >= 0), treatment = group),
    aes(x, ymin = ymin, ymax = ymax, group = group),
    fill = "red",
    alpha = 0.2) +
  xlab("Subject Updating (Posterior - Prior)") +
  ylab(NULL) +
  xlim(-20, 20)
```

```
## Picking joint bandwidth of 1.25
```



```
#ggsave("density.pdf", width = 8, height = 6)
```

it is hard to provide the summary information that goes into a density plot. But the mean and standard deviations is well described using the summary function applied to the differences.

```
summarySE(complete, measurevar="difference", groupvars=c("treatment"), na.rm = TRUE)
```

```
## treatment N      mean      sd      se      ci      name
## 1   Amami  97 -1.5979381  4.702968  0.4775141  0.6161997 difference
## 2 Liancourt 41  0.2195122  5.547577  0.8663860  1.1289677 difference
## 3   Senkaku 47 -0.9361702  4.659746  0.6796937  0.8837569 difference
```

## **C Abstract experiment**

We now include the documentation for the abstract experiment. For a discussion of how the experiment fits APSA principals with human subjects see the section above.

### **C.1 Qualtrics print out**

## **consent**

### **Consent Form.**

This is a study meant to gauge public opinion on issues of U.S. foreign policy including military conflict. My name is [REDACTED] and I am a researcher at [REDACTED]. Participation in this study will involve completing the following survey. The survey will take approximately 10 minutes.

There are no known or anticipated risks to you for participating in the survey. The benefit to you comes in the form of the compensation you were offered when you were recruited into the survey. Only respondents who correctly answer attention check questions will be allowed to complete the survey. Failure to pass this check may affect compensation. In addition, we hope that this survey can also lead to a better understanding about important issues U.S. politicians must consider. All of your responses will be anonymous.

Only the researchers involved in this study and those responsible for research oversight will have access to the

information you provide. Your responses will be recorded in a dataset. No identifying information will be collected or stored in our dataset.

Participation in this study is completely voluntary. You are free to decline to participate, to end participation at any time for any reason, or to refuse to answer any individual question. Your decision whether or not to participate in this study will not affect your relationship with [REDACTED] or LUCID.

If you have any questions about this study, you may contact the researchers [REDACTED] [REDACTED] if you would like to talk with someone other than the researcher to discuss problems or concerns, to discuss situations in the event that the researcher is not available, or to discuss your rights as a research participant, you may contact th [REDACTED]

Are you willing to participate in this study?

- Yes
- No

## **prompt**

We will describe a hypothetical foreign policy scenario. It focuses on a fictional country named Bandaria.

We will:

1. Present information about Bandaria's foreign policy
2. Ask you to evaluate Bandaria's strategic intentions.

Similar hypothetical scenarios are used to form US foreign policy. Please take it seriously.

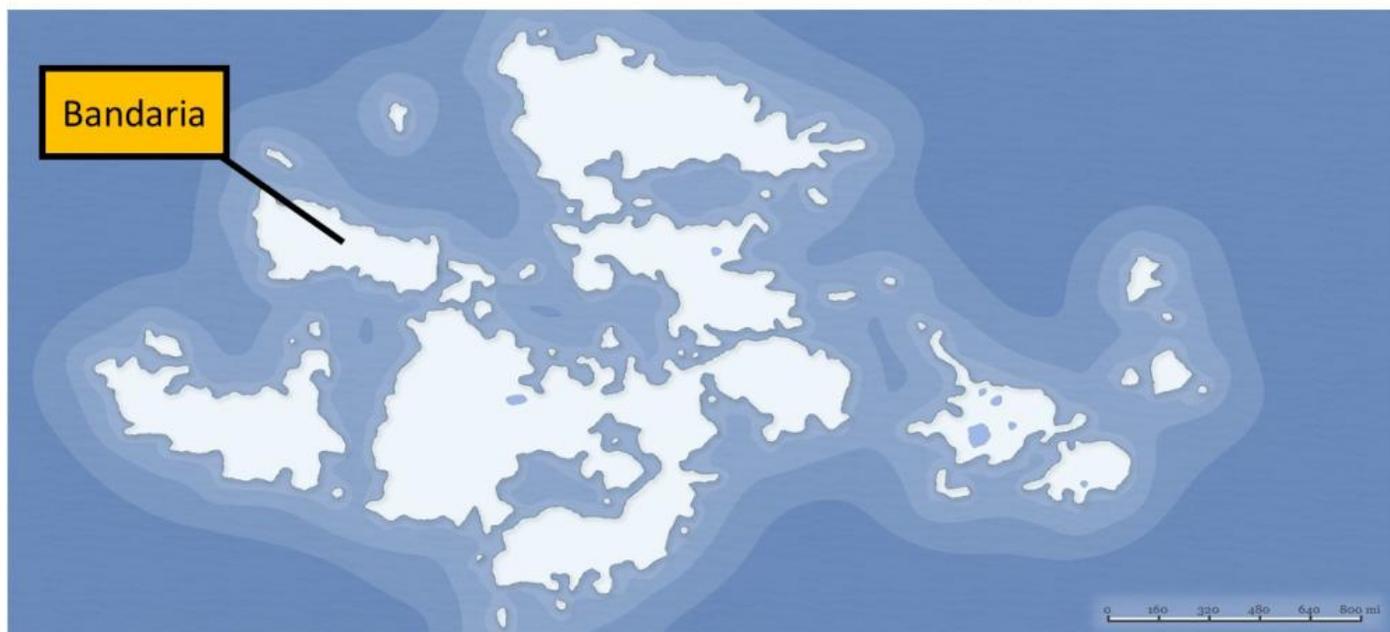
## **factual**

Bandaria is 1 of 5 countries in its region.

Over the centuries these countries have fought wars with each other over the small islands between them.

Bandaria's current territorial borders consist of one large Island. But its neighbors control many smaller Islands.

## Bandaria and surrounding Islands



What is the name of the main country in this scenario

- Arcadia
- Bandaria
- India
- Country A

## **prior**

In the last 10 years Bandaria has experienced enormous economic and military growth.

For the first time ever, Bandaria is the most powerful country in its region. If it wants to, it can easily take territory through force.

The question is: does Bandaria want to use its military to take territory? If it does, how much territory does it want to take?

What are Bandaria's territorial objectives? There are two most likely answers.

Bandaria's government claims its territorial ambitions are limited to small Islands of historical importance. It states:

“Several Islands of vital historical significance were stolen from us. We will re-take small islands of historical significance, through force if necessary. However, we hold no other territorial ambitions.”

A neighboring country provided an alternative assessment:

“Bandaria’s appeal to limited aims is a cover for its true ambition: regional domination. Bandaria is an opportunist that will selectively expand into territories when the economic and strategic value outweighs the costs of fighting. If we do not unite against Bandaria now, Bandaria will take territories of strategic value and dominate our region over time.”

We just described two kinds of territorial ambitions that Bandaria could hold. Pick the two that best fit what we described.

- Bandaria will not seek any territorial revision
- Bandaria's territorial ambition is limited to small islands of historical importance
- Bandaria seeks regional domination.

**prior\_response**

Correct!

We now want your opinion about how likely these outcomes are relative to each other. In summary, here are the two possibilities:

A. **Limited**: Bandaria will at most expand its borders to take small islands of historical value.

B. **Expansive**: Bandaria will opportunistically expand its borders across its region. Bandaria may not always fight. But it will expand when the strategic and economic value outweighs the cost.

How likely are these two outcomes relative to each other?



**Historical\_value\_fight**

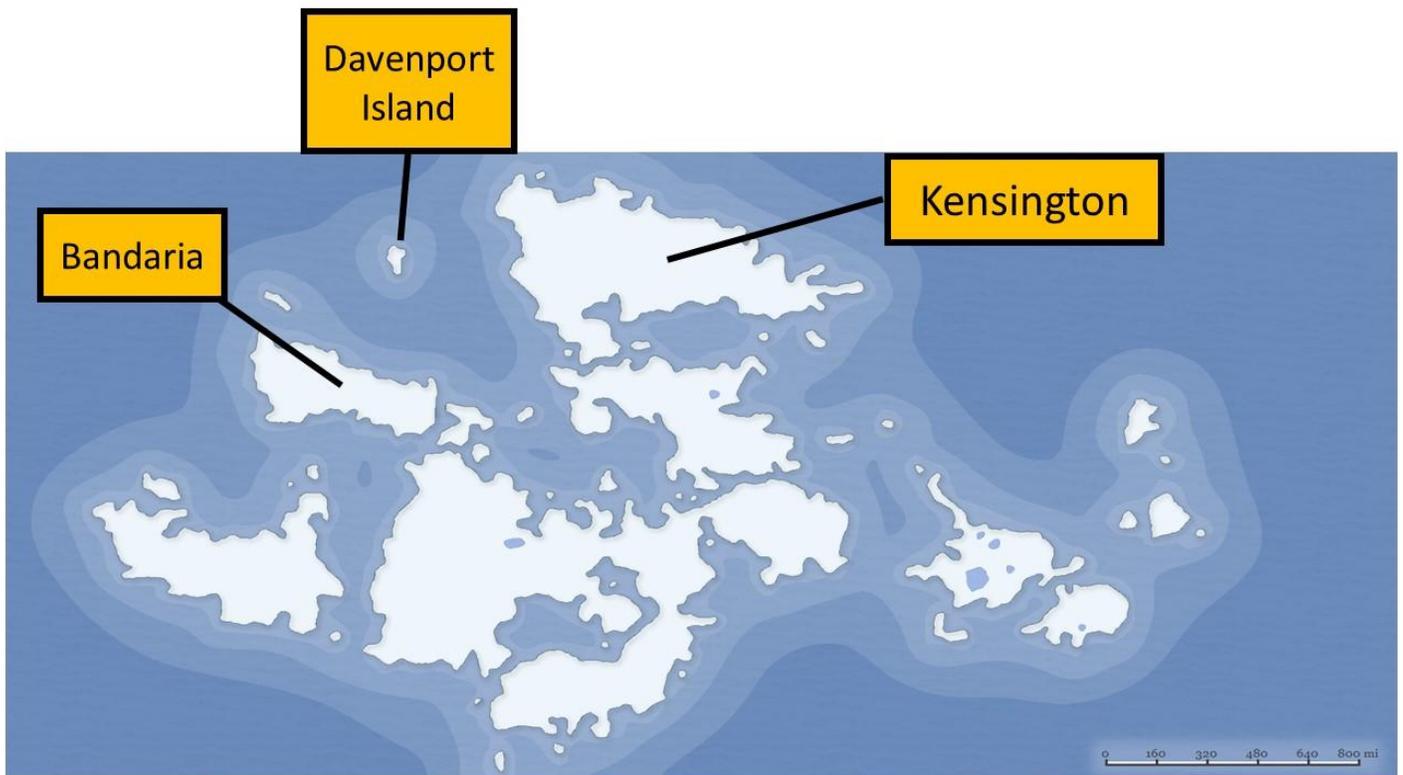
**A crisis over the Davenport Island**

Kensington is one of Bandaria's neighbors.

Bandaria just learned that Kensington plans to build 10,000 homes on the Davenport Island and populate the Davenport Island with Kensington nationals.

Bandaria is easily strong enough to take the Davenport Island if it acts now. But it will lose many ships it would need to complete other military missions.

If Bandaria waits, Kensington will fortify the Davenport Island. Bandaria will not be able to get it back for decades.



For Bandaria, the Davenport Island holds large historical value but no economic or strategic value.

Bandaria controlled the Davenport Island for 1000 years until Kensington invaded in 1905. However, the Davenport Island contains no natural resources and is not well situated to help Bandaria expand into neighboring countries.

Weighing these costs and benefits, Bandaria **invaded** the Davenport Island.

As expected, Bandaria took this Island of vital historical importance to Bandaria but no strategic value. Bandaria lost military assets it would need to complete other military missions.

# Historical\_nofight

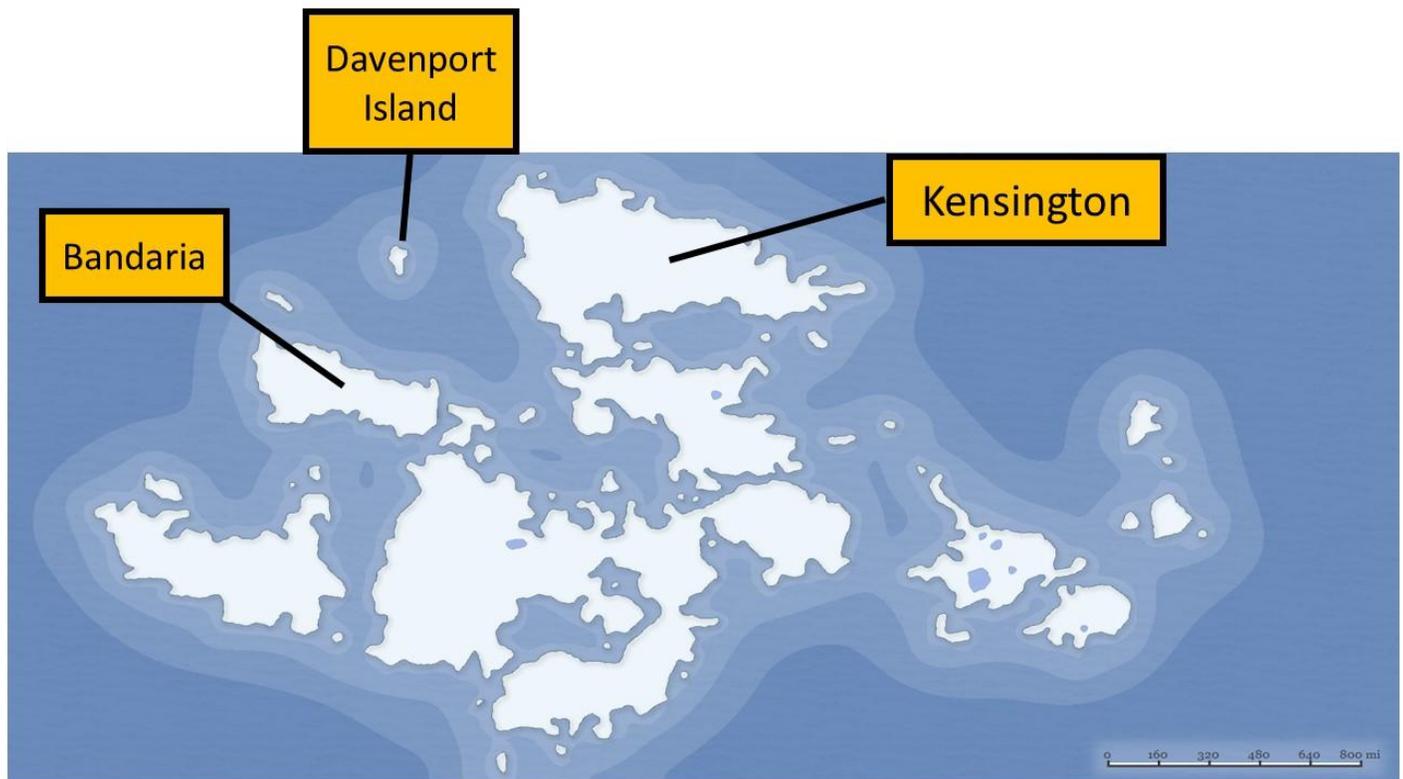
## A crisis over the Davenport Island

Kensington is one of Bandaria's neighbors.

Bandaria just learned that Kensington plans to build 10,000 homes on the Davenport Island, and populate the Davenport Island with Kensington nationals.

Bandaria is easily strong enough to take the Davenport Island if it acts now. But Bandaria will lose many ships it would need to complete other military missions.

If Bandaria waits, Kensington will fortify the Davenport Island. Bandaria will not be able to get it back for decades.



For Bandaria, the Davenport Island holds large historical value but no economic or strategic value.

Bandaria controlled the Davenport Island for 1000 years until Kensington invaded in 1905. However, the Davenport Island contains no natural resources and is not well situated to help Bandaria expand into neighboring countries.

Weighing these costs and benefits, Bandaria **did not** invade the Davenport Island.

As expected, Kensington consolidated power over this island of vital historical importance to Bandaria but no strategic value. Now, Bandaria will not be able to take the Davenport Island back for decades.

Instead, Bandaria kept military assets it needs to complete other military missions.

## **Strategic\_values**

### **A crisis over the Davenport Island**

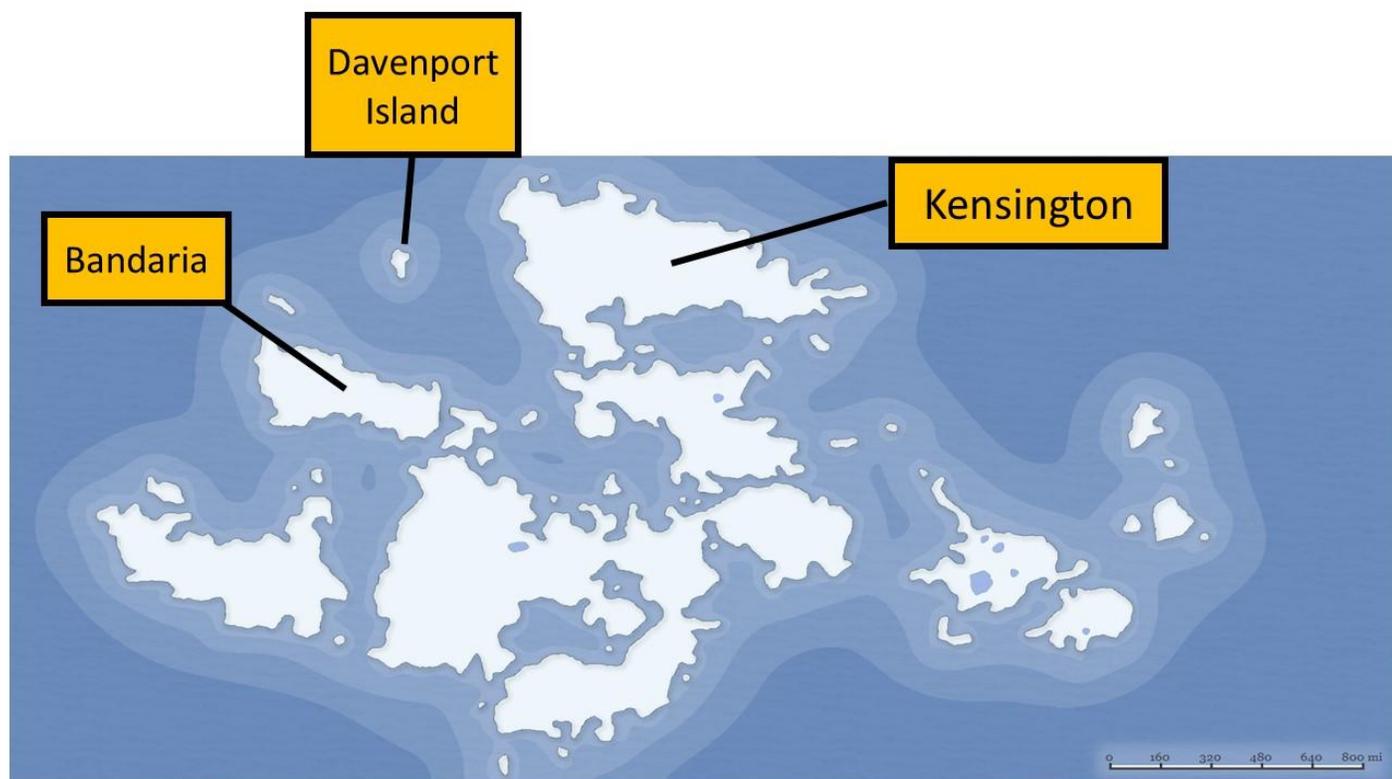
Kensington is one of Bandaria's neighbors.

Bandaria just learned that Kensington plans to build 10,000 homes on the Davenport Island, and populate the Davenport Island with Kensington nationals.

Bandaria is easily strong enough to take the Davenport Island if it acts now. But Bandaria will lose many ships if

would need to complete other military missions.

If Bandaria waits, Kensington will fortify the Davenport Island. Bandaria will not be able to get it back for decades.



For Bandaria, the Davenport Island holds no historical value but huge economic and strategic value.

Bandaria has never controlled the Davenport Island. However, the Davenport Island sits on the largest oil deposit in the region. It is also well situated as a strategic choke-

point to cut off neighboring countries from other parts of the world.

Weighing these costs and benefits, Bandaria invaded the Davenport Island.

As expected, Bandaria took this Island of little historical importance but vital economic and strategic importance. Bandaria lost military assets it would need to complete other military missions.

## **post\_response**

With this new information, we want you to re-assess the relative likelihood of the following two possibilities:

A. **Limited**: Bandaria will at most expand its borders to take Islands of historical value.

B. **Expansive**: Bandaria will opportunistically expand its borders across its region. Bandaria may not always fight.

But it will expand when the strategic and economic value outweighs the cost.

How likely are these two outcomes relative to each other?



Did the crisis event influence your assessment? If it did, why?

Which option best summarizes Bandaria's crisis response:

- Invaded Island
- Did not invade Island

Was the Davenport Island important to Bandaria for historical reasons?

- yes
- no

## **ps\_1**

To what extent do you agree or disagree with the following question:

States with limited aims care more about issues of historical and cultural importance than states with expansive interests.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Powered by Qualtrics

## **C.2 R markdown for analysis of results**

# Abstract Design

In what follows, we generate the analytic tests from our abstract vignete. You'll notice that the code and standards for inference are the same as the main experiment for H1 and H2. H3 (which is new) exactly extends H2.

A copy of the qualtrics print-out of the instrument is available in the next sub-section.

Our packages:

```
require(tidyverse)

## Loading required package: tidyverse
## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5    v purrr  0.3.4
## v tibble  3.1.3    v dplyr  1.0.7
## v tidyr   1.1.3    v stringr 1.4.0
## v readr   2.0.1    v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
require(qualtRics)

## Loading required package: qualtRics
## Warning: package 'qualtRics' was built under R version 4.1.2
#library("gridExtra")
library(coin)

## Warning: package 'coin' was built under R version 4.1.2
## Loading required package: survival
library(plyr)

## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
```

```
## The following object is masked from 'package:purrr':
##
## compact
library(ggribes)
```

## Downloading data

```
readRenviron("~/Renviron")
surveys <- all_surveys()
response <- fetch_survey(surveyID = surveys$id[1], verbose = TRUE, convert = TRUE, label = TRUE, force = TRUE)

## | |
##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   StartDate = col_datetime(format = ""),
##   EndDate = col_datetime(format = ""),
##   Status = col_character(),
##   IPAddress = col_character(),
##   Finished = col_logical(),
##   RecordedDate = col_datetime(format = ""),
##   ResponseId = col_character(),
##   RecipientLastName = col_logical(),
##   RecipientFirstName = col_logical(),
##   RecipientEmail = col_logical(),
##   ExternalReference = col_logical(),
##   DistributionChannel = col_character(),
##   UserLanguage = col_character(),
##   consent_yn = col_character(),
##   attention_1 = col_character(),
##   attention_2_1 = col_character(),
##   attention_2_2 = col_character(),
##   attention_2_3 = col_character(),
##   `Q108_First Click` = col_logical(),
##   `Q108_Last Click` = col_logical()
##   # ... with 8 more columns
## )
## i Use `spec()` for the full column specifications.
## Warning in file.remove(survey.fpath): cannot remove file 'C:/Users/micha/
## AppData/Local/Temp/RtmpW0m9Gf/Violent_trust_abstract_launch.csv', reason
## 'Permission denied'
```

## Cleaning data

This involves defining the treatment variable, and removing subjects that failed attention checks. Consistent with the main experiment, I remove subjects that are inattentive.

```
response$treatment <- NA
response$treatment[which(response$FL_32_DO_Historical_nofight==1)] <- "Historical, doesn't fight"
response$treatment[which(response$FL_32_DO_Historical_value_fight==1)] <- "Historical, fights"
response$treatment[which(response$FL_32_DO_Strategic_values==1)] <- "Strategic, fights"
```

```
response$difference <- response$post_response_1 - response$prior_response_1
complete <- response %>% filter(is.na(response$treatment)==F)
```

We now have a summary of the distribution of treatment.

```
table(response$treatment)
```

```
##
## Historical, doesn't fight      Historical, fights      Strategic, fights
##                               37                          34                          35
```

## Analysis

### Hypothesis 1

Consistent with the main study, hypothesis 1 considers those who observe Bandaria fight for an issue with high historical value. We focus on the difference in their responses pre and post treatment. Our inference is based on the p-value derived from a permutation test that examines the difference in means. We define our standards for inference as follows:

p-value < .05 I infer support p-value < .01 I infer strong support p-value < .1 I infer moderate support

Here is our code, identical in the main study:

```
complete %>% filter(treatment=="Historical, fights") %>%
  select(treatment,post_response_1,prior_response_1) %>%
  pivot_longer(cols = post_response_1:prior_response_1,
               names_to = "timing",
               values_to = "value") %>%
  mutate(timing = as.factor(timing)) %>%
  independence_test(value ~ timing, data = .,
                   ytrafo = rank_trafo, distribution = exact(), alternative="less")
```

```
##
## Exact General Independence Test
##
## data: value by
## timing (post_response_1, prior_response_1)
## Z = -2.1379, p-value = 0.01608
## alternative hypothesis: less
```

As these results show, we find support for H1.

### Hypothesis 2

Consistent with the main study, hypothesis 2 considers those who observe Bandaria fight for an issue with high historical value against those who observe Bandaria fight for an issue with high strategic value. Our inference is based on the p-value derived from a permutation test that examines the difference in means in these groups. We define our standards for inference as follows:

p-value < .05 I infer support p-value < .01 I infer strong support p-value < .1 I infer moderate support

Here is our code, identical in the main study:

```
complete %>%
  filter(treatment=="Historical, fights"|treatment=="Strategic, fights") %>%
  mutate(treatment = as.factor(treatment)) %>%
```

```
independence_test(post_response_1 ~ treatment,
  data = ., ytrafo = rank_trafo, distribution = exact(), alternative="less")
```

```
##
## Exact General Independence Test
##
## data: post_response_1 by
## treatment (Historical, fights, Strategic, fights)
## Z = -3.1983, p-value = 0.0005766
## alternative hypothesis: less
```

As you can see, we infer strong support.

### Hypothesis 3

Like hypothesis 2, hypothesis 3 contrasts two post-treatment groups. Both groups observe a crisis over an issue of high historical value. In one group Bandaria fights, in the other Bandaria does not. H3 states that fighting leads to a higher level of trust. We use the same design as H2. Our inference is based on the p-value derived from a permutation test that examines the difference in means in these groups. We define our standards for inference as follows:

p-value < .05 I infer support p-value < .01 I infer strong support p-value < .1 I infer moderate support

Here is our code, identical in the main study:

```
complete %>%
  filter(treatment=="Historical, fights"|treatment=="Historical, doesn't fight") %>% mutate(treatment)
  independence_test(post_response_1 ~ treatment, data = .,
    ytrafo = rank_trafo, distribution = exact(), alternative="greater")
```

```
##
## Exact General Independence Test
##
## data: post_response_1 by
## treatment (Historical, doesn't fight, Historical, fights)
## Z = 2.2813, p-value = 0.01102
## alternative hypothesis: greater
```

As you can see, we infer strong support.

### Visualizing Results

We use the same function that we used in the main experiment

```
summarySE <- function(data=NULL, measurevar, groupvars=NULL, na.rm=FALSE,
  conf.interval=.80, .drop=TRUE) {
  library(plyr)

  # New version of length which can handle NA's: if na.rm==T, don't count them
  length2 <- function(x, na.rm=FALSE) {
    if (na.rm) sum(!is.na(x))
    else      length(x)
  }

  # This does the summary. For each group's data frame, return a vector with
  # N, mean, and sd
```

```

datac <- ddply(data, groupvars, .drop=.drop,
  .fun = function(xx, col) {
    c(N = length2(xx[[col]], na.rm=na.rm),
      mean = mean (xx[[col]], na.rm=na.rm),
      sd = sd (xx[[col]], na.rm=na.rm)
    )
  },
  measurevar
)
datac$se <- datac$sd / sqrt(datac$N) # Calculate standard error
ciMult <- qt(conf.interval/2 + .5, datac$N-1)
datac$ci <- datac$se * ciMult
datac$name <- as.factor(measurevar)
return(datac)
}

```

We then compute the summary stats

```

su.data <- summarySE(complete, measurevar="post_response_1", groupvars=c("treatment"), na.rm = TRUE)

su.data$low.ci <- su.data$mean - su.data$ci
su.data$high.ci <- su.data$mean + su.data$ci
su.data$treatment <- as.factor(su.data$treatment)

```

The datatable for the ladder plot is

```

su.data

##           treatment  N    mean    sd      se      ci
## 1 Historical, doesn't fight 37 3.4054054 5.035758 0.8278735 1.0808004
## 2 Historical, fights 33 0.1212121 5.951668 1.0360524 1.3557499
## 3 Strategic, fights 35 4.7714286 4.263978 0.7207439 0.9419774
##           name    low.ci  high.ci
## 1 post_response_1  2.324605 4.486206
## 2 post_response_1 -1.234538 1.476962
## 3 post_response_1  3.829451 5.713406

```

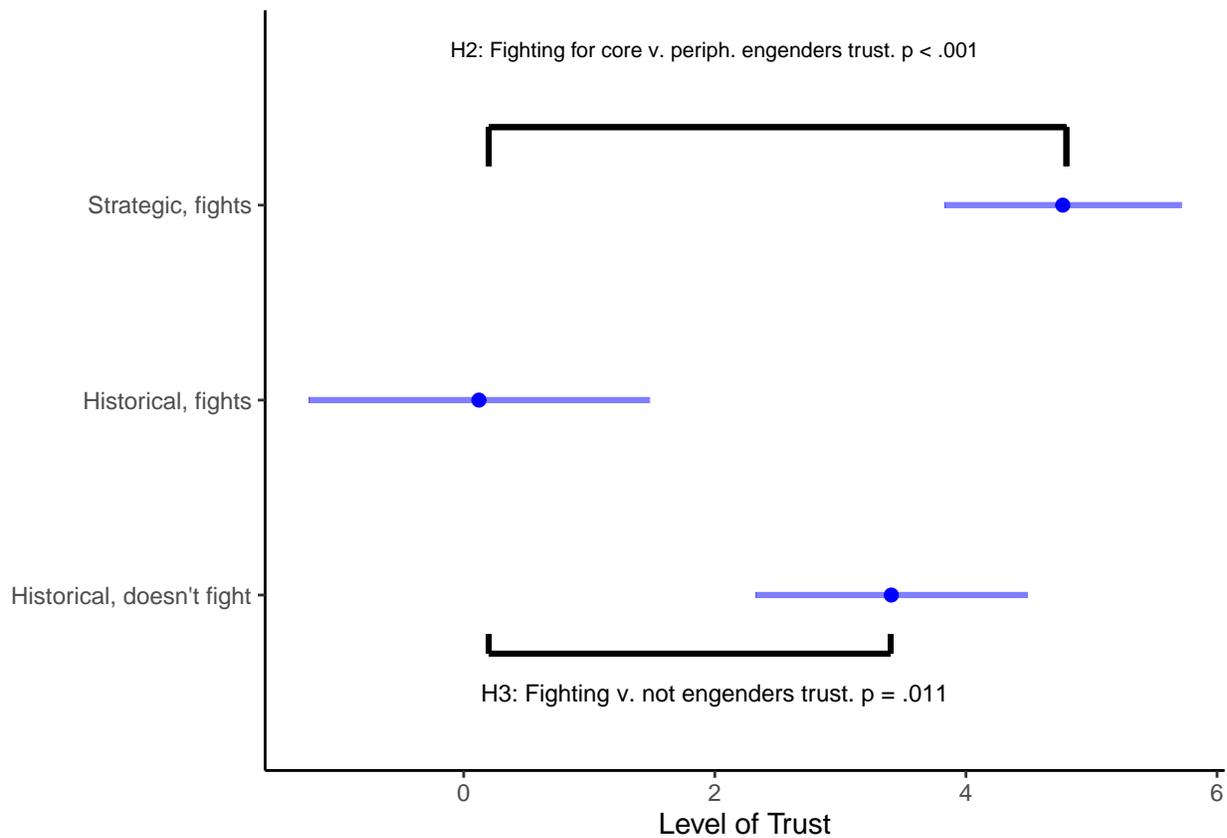
Here is the plot

```

ggplot(su.data, aes(x=mean, y=treatment) ) +
  geom_point(size = 2 , position = position_dodge(width = 0.3), color="blue") +
  geom_errorbarh(aes(xmin = low.ci , xmax = high.ci, height = 0),
size=1.1, alpha=.5, position = position_dodge(width = 0.3), color="blue") +
  theme_classic() +
  theme(legend.position = "NULL") + labs(x = "Level of Trust", title = NULL, y=NULL) +
  geom_segment(size = 1, x=.2,y=0.7, xend= 0.2 , yend=.8) +
  geom_segment(size = 1, x=3.4,y=.7, xend= 3.4 , yend=.8) +
  geom_segment(size = 1, x=.2,y=.7, xend= 3.4 , yend=.7)+
  annotate("text", x = 2, y = .1, label = "") +
  annotate("text", x = 2, y = .5,
size = 3.2, label = "H3: Fighting v. not engenders trust. p = .011") +
  geom_segment(size = 1, x=.2,y=3.2, xend= 0.2 , yend=3.4) +
  geom_segment(size = 1, x=4.8,y=3.2, xend= 4.8 , yend=3.4) +
  geom_segment(size = 1, x=.2,y= 3.4, xend= 4.8 , yend=3.4) +
  annotate("text", x = 2, y = 4, label = "") +
  annotate("text", x = 2, y = 3.8, size = 2.8,

```

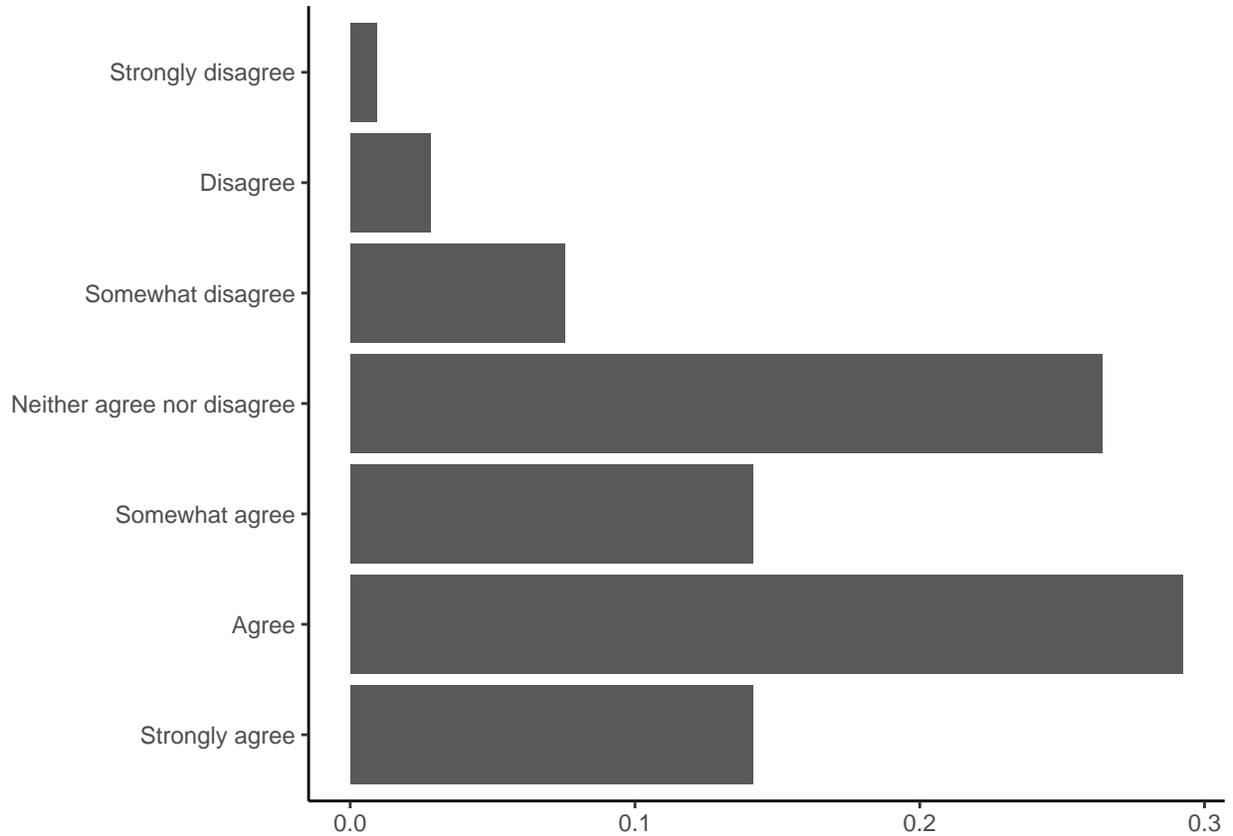
```
label = "H2: Fighting for core v. periph. engenders trust. p < .001")
```



```
#ggsave("abst_ladder.pdf", width = 8, height = 4)
```

We also include a histogram of data on a post-survey question.

```
ploter <- as.data.frame(prop.table(table((complete$limited_direct_2))))  
ploter <- ploter %>% filter(Var1 != "<NA>") # Removes NAs  
  
ggplot(ploter, aes(Var1, Freq)) +  
  geom_bar(stat = "identity") + coord_flip() + xlab(NULL) + ylab(NULL) +  
  theme_classic()
```



```
#ggsave("abst_hist.pdf", width = 8, height = 4)
#complete$post_text
```

The data that makes up that visualization is:

```
table(complete$limited_direct_2) # As a count
```

```
##
##          Strongly agree          Agree
##          15                    31
##          Somewhat agree Neither agree nor disagree
##          15                    28
##          Somewhat disagree          Disagree
##          8                        3
##          Strongly disagree          <NA>
##          1                        5
```

```
prop.table(table(complete$limited_direct_2)) # as a proportion.
```

```
##
##          Strongly agree          Agree
##          0.141509434          0.292452830
##          Somewhat agree Neither agree nor disagree
##          0.141509434          0.264150943
##          Somewhat disagree          Disagree
##          0.075471698          0.028301887
##          Strongly disagree          <NA>
##          0.009433962          0.047169811
```