

Knowing Intentions: What Do You See—Actors or Actions?

Seok Joon Kim

Department of Political Science and Diplomacy

Kyungpook National University

sjkim.ir@gmail.com

Abstract

Do ordinary citizens place more credibility on signals from democracies than non-democracies? When and under what conditions do they assess democratic signals as more or less credible than those from non-democracies? This paper argues that democracies can make their signals more credible than non-democracies and can be perceived as such but offers a new explanation for why democratic credibility can *still* go unobserved. This study examines democratic advantage in signaling by decomposing it into two components: the perceived signaling capability gap—the difference in how observers evaluate the signaling capabilities of democracies versus non-democracies, and the prior belief difference—the difference in *prior beliefs* about the inherent aggressiveness of both regime types. This model offers a robust theoretical framework for understanding how prior beliefs about a signaler's aggressiveness can independently affect the credibility of state signals during crises. This model is tested through survey experiments in the United States and China.

Keywords: democratic credibility, democratic perception model, prior beliefs.

Do ordinary citizens place more credibility on state signals from democracies than non-democracies? When and under what conditions do observers of state actions assess democratic signals as more or less credible than those from non-democracies? Over the past three decades, IR scholars have extensively debated Fearon's "plausible working hypothesis" of democratic credibility (Fearon 1994, 582) that democratic institutions make democracies more effective at signaling their resolve for war to foreign audiences compared to non-democracies (Fearon 1994; Schultz 1998; Smith 1998; Guisinger and Smith 2002; Ramsay 2004; Tomz 2007; Fang 2008; Kurizaki and Whang 2015).

Critics have struggled to find empirical evidence supporting the democratic advantage in signaling during crises (Slantchev 2006; Weeks 2008; Trager 2010; Snyder and Borghard 2011; Trachtenberg 2012; Downes and Sechser 2012). This holds true even in controlled experimental settings (Kertzer et al., 2021, 308, 319-24; Renshon et al., 2023). Critics of democratic credibility argue that a democratic audience might prefer empty threats that would bring better policy outcomes over implementing threats that would result in policy failures (Snyder and Borghard 2011; Trachtenberg 2012; Levendusky and Horowitz 2012; Nomikos and Sambanis 2019). In a similar vein, non-democracies may signal their intentions effectively through mechanisms such as military actions (Slantchev 2005), internal political punishment (Brown and Marcum 2011), concerns about reputations (Sartori 2002; Li and Chen 2021), and other methods.

Contrary to the literature that rejects democratic credibility in signaling, this paper argues that democracies can make their signals more credible than non-democracies and can be perceived as such but offers a new explanation for why democratic credibility can *still* go unobserved by the foreign public. It distinguishes between observers' perceptions of a state's action in conjunction with their perceptions of the state's intrinsic nature as an actor. Specifically, it examines democratic advantage in signaling by decomposing it into two components: the perceived signaling capability gap—the difference in how observers evaluate the signaling capabilities of democracies versus non-democracies, and the prior belief difference—the difference in *prior beliefs* about the inherent aggressiveness of both regime types. This model offers a robust theoretical framework for understanding how prior beliefs about a signaler's aggressiveness can

independently affect the credibility of state signals during crises. This model is tested through survey experiments in the United States and China.

This research focuses on the psychological dynamics of ordinary citizens' assessments of democratic credibility rather than that of elites. The question of whether public opinion influences policymakers' foreign policy preferences has long been debated (For example, see Converse 2006, Guisinger and Elizabeth 2017, Aldrich, Sullivan, and Borgida 1989; Datta 2009; Holsti 1992; Goldsmith and Horiuchi 2012). While this issue is beyond the scope of this paper, the study of the cognitive psychology of the mass public has recently gained prominence in IR, largely due to contributions from scholars of human political behavior who suggest that public opinion shapes politicians' foreign policies primarily because of the impact on election outcomes (for example, see Yarhi-Milo, Kertzer, and Renshon 2018; Tomz, Weeks, and Yarhi-Milo 2020).

The experiments conducted to test the plausibility of the proposed model yielded several key findings. First, US citizens commonly view democracies' signaling capabilities as superior to that of non-democracies, a perception only partly shared by Chinese citizens. Second, holding other factors constant, US and Chinese citizens hold different prior beliefs about democracies and non-democracies regarding their inherent peacefulness or aggressiveness. US citizens tend to view democracies as more peaceful and less threatening than non-democracies, while Chinese citizens generally have a more negative perception of democracies. Third, these distinct evaluations of signaling capabilities and different prior beliefs about signaler aggressiveness led to variations in the visibility of the democratic advantage at signaling depending on the type of signals among US and Chinese citizens. US citizens' positive pre-existing beliefs towards democracy bolster the credibility of a democracy's benign signals but reduce the credibility of its hostile signals. Particularly, the observed democratic credibility was most pronounced among US citizens when state intentions were conveyed through diplomatic statements rather than military actions. Conversely, the Chinese citizens' negative pre-existing beliefs towards democracy tend to undermine the credibility of a democracy's benign signals, while their beliefs do not have any significant impact on the credibility of its hostile signals.

This study makes significant implications in the study of democratic credibility and democratic peace. First, it clarifies the concept of democratic credibility in signaling while providing a more comprehensive framework by extending the theory of democratic credibility

beyond its traditional focus on resolve for war by incorporating signals of benign intentions. Second, more broadly, this paper offers insights into the micro-level mechanisms of institutional and constructivist approaches to democratic peace. It shows how the institutional features of democracies make their signals credible, while illustrating how prior beliefs, perhaps socially constructed, about a regime's peaceful or aggressive nature independently affect democratic credibility, and in turn, democratic peace as well.

Democratic Credibility Theory and its Critics

States often want to signal their intentions to audiences in target countries during crises. The credibility of these signals has long been debated among IR scholars. Since Fearon (1994), prominent IR signaling theorists have proposed that democracies are better equipped than non-democracies at conveying their resolve for war. They argue that certain features of domestic political institutions give democracies an advantage in signaling their intentions—what is now known as democratic credibility theory (Fearon 1994; Schultz 1998; Smith 1998; Guisinger and Smith 2002; Ramsay 2004; Fang 2008). For instance, Fearon suggests that if a democratic leader escalates a crisis and then backs down, they risk electoral backlash for tarnishing the national reputation and pride. This potential for domestic political cost bolsters the credibility of a democracy's public threats (Fearon 1994; Uzonyi, Souva, and Golder 2012; McManus 2014.). Schultz (1998) also suggests a democratic advantage in signaling through domestic political competition. The support or opposition of an opposition party to a government's decision to use military force can serve as a secondary source of credibility. Moreover, the presence of an opposition party prompts the government to be more selective about the threats it issues. Although they differed on the exact mechanisms, early IR signaling theorists typically agreed that democratic signalers had a *superior signaling capability* to update observers' beliefs.¹

¹ Note that the discussion of various mechanisms through which democracies communicate their intentions better than non-democracies, such as belligerence and incompetence (Uzonyi, Souva, and Golder 2012; Kertzer and Brutger 2016; Nomikos and Sambanis 2019), is beyond the scope of this paper.

Critics of democratic credibility theory have challenged its principles on both theoretical and empirical grounds. First, they contend that non-democratic leaders can also generate audience costs. For example, Weeks (2008) asserts that leaders of autocracies—specifically those in single-party or military regimes—can signal their resolve to observers as effectively as democratic leaders. Weiss (2013) advances a similar argument, maintaining that authoritarian leaders can signal their resolve to foreign publics by either permitting or suppressing anti-foreign protests, as handling nationalistic protests is risky and costly in authoritarian states; thus, their responses to such demonstrations provide information about the government’s foreign policy interests. Brown and Marcum (2011, 142–43) take this argument further, contending that autocratic leaders may be more accountable to their selectorate than democratic ones, as dictatorships typically have smaller winning coalitions with more effective mechanisms for monitoring and sanctioning leaders.

Second, certain signaling theorists reject democratic credibility theory, claiming that non-democracies can signal their resolve as democracies via military mobilization (Slantchev 2005), by upholding a reputation for honesty (Sartori 2002; 2005; Li and Chen 2021), or by risking adverse reactions from target countries (Trager 2010). Slantchev posits that autocracies can use military means to signal their resolve as effectively as democracies (Slantchev 2005). Sartori (2002) also contends that even in the absence of domestic audience costs, diplomacy is effective due to concerns over maintaining an international reputation for honesty. Trager (2010) suggests that signals from non-democracies can be assessed as credible as those from democracies because signalers should consider the potential for provoking unfavorable reactions from the target country.

Third, several IR scholars dismiss democratic credibility theory based on empirical evidence and historical data (Snyder and Borghard 2011; Downes and Sechser 2012; Trachtenberg 2012; Yarhi-Milo, Kertzer, and Renshon 2018). Downes and Sechser (2012) critique prior research favoring democratic credibility theory, claiming that coercive threats issued by democracies were not as successful as previously claimed. Similarly, many IR scholars, such as Kertzer and his colleagues, failed to find empirical support for the democratic advantage in their survey experiments, finding that neither ordinary citizens nor political elites perceived a democratic advantage in signaling (Kertzer, Renshon, and Yarhi-Milo 2021, 308, 319–24; Renshon, Yarhi-Milo, and Kertzer 2023).

State Intentions

In international relations, state intentions are defined as “the actions he [an actor] will take under given circumstances” (Jervis 1976, 48) or refer to a state’s foreign policy plans concerning the status quo (Yarhi-Milo 2013, 10). These are often utilized in two distinct IR contexts: the security dilemma and crisis bargaining. While scholars in both areas share an interest in deciphering how states interpret intentions, their focuses diverge. Security dilemma scholars concentrate on how *benign* (status quo) states, as opposed to *greedy* (revisionist) states, reveal peaceful intentions to escape the security dilemma (Glaser 1997; Kydd 1997; 2007). Conversely, crisis bargaining scholars focus on how *resolved* states, as opposed to *unresolved* states, signal their willingness to fight to deter adversaries (Fearon 1994; 1997). As states strive to convey their intentions, those involved in security dilemmas or crisis bargaining are equally eager to assess the credibility of these signals. To be considered credible, benign signals should be interpreted as peaceful and reassuring, while hostile signals should be perceived as aggressive and threatening. Benign and unresolved states are less likely to send hostile signals to avoid conflict, while aggressive and resolved are more likely to issue threats and engage in conflict.

Words and Deeds

IR scholars propose that states embroiled in disputes or crises communicate their intentions through various means, typically diplomatic statements and military actions. They have often debated which types of signaling are more effective (Fearon 1997; Slantchev 2005; Fuhrmann and Sechser 2014; McManus 2014; Yarhi-Milo, Kertzer, and Renshon 2018; Katagiri and Min 2019). When analyzing the effectiveness of signals, it is important to distinguish between different signaling modes. Without this distinction, analyzing aggregated signals can obscure the effectiveness of one type of signaling mode due to the influence of another type. There are reasons to suspect that the effects of different signal types will not be the same across all regime types. Fearon (1997, 68, 82) emphasizes the potency of democratic leaders’ words in revealing state intentions, arguing that the domestic political cost incurred by democratic leaders when bluffing lends greater credibility to their words compared to those of non-democratic counterparts. Beyond verbal statements, states can also demonstrate their intentions through military actions such as

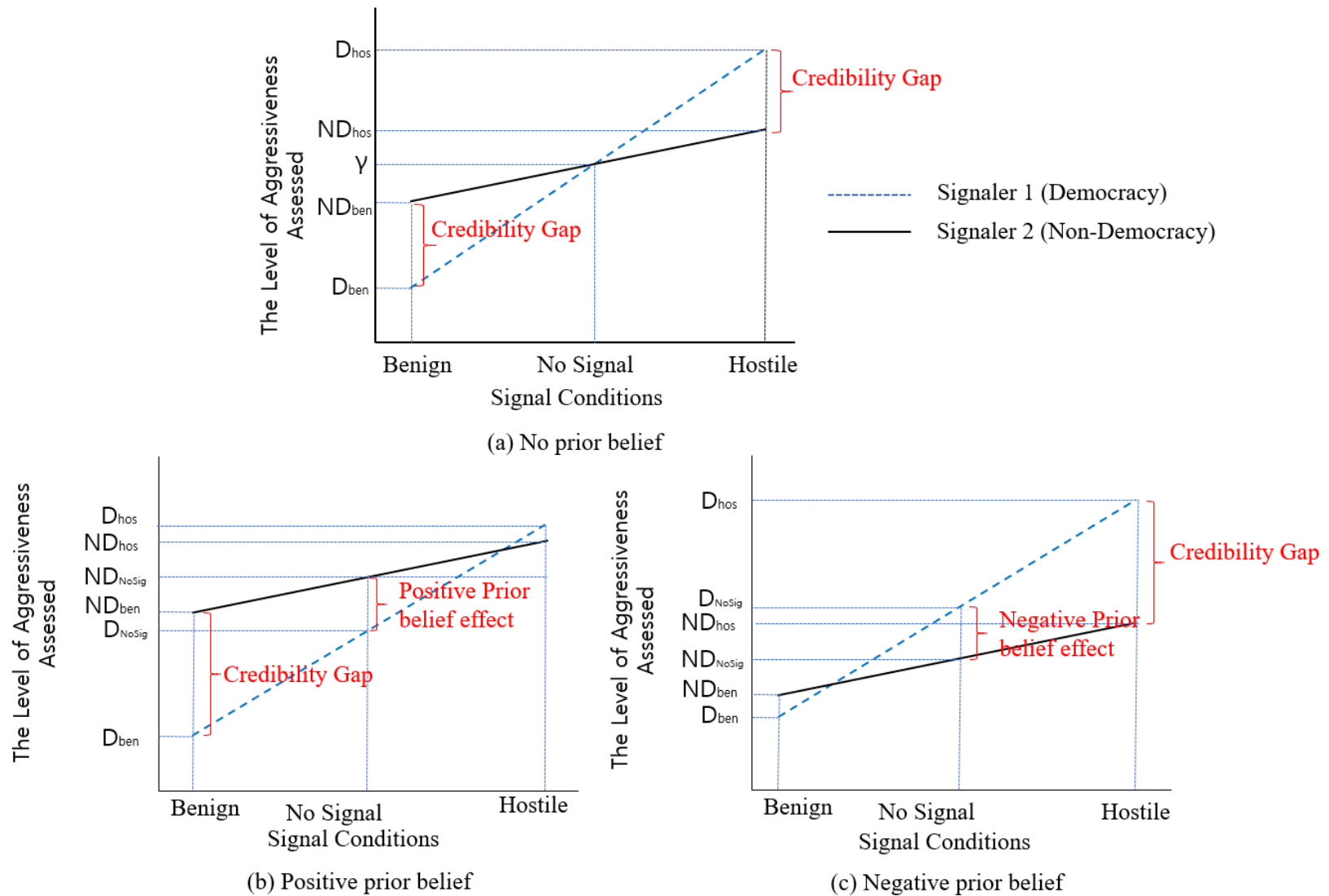
military drawdowns or mobilizations (Glaser 2010; Slantchev 2005). Fearon (1994, 83) speculated that signaling via military mobilization might play a more significant role in authoritarian countries than in democratic ones. Similarly, Slantchev (2005, 546) claimed, "...it does not necessarily follow that democracies would be able to signal their resolve any better in a crisis in which military means are available to autocracies as well." Thus, while the words of democratic leaders may be perceived as more credible, the perceived credibility of military actions may be comparable between democracies and non-democracies (Slantchev 2005). This suggests that debates over democratic credibility theory will benefit from a granular approach that disaggregates the types of signals and the reasons for their credibility.

Democratic Perception Model

As noted before in this paper, theorists of democratic credibility believe that democracies' signals more credibly reveal their intentions than those of non-democracies due to their *superior signaling capability*. However, most IR scholars studying democratic credibility have focused on coercive threats, often overlooking reassurances. This gap leaves questions about whether the logic of democratic credibility in threats applies similarly to reassurances. If democratic credibility is as real as some theorists claim, democracies might also excel in signaling benign intentions. For instance, democratic leaders who promise peace but then attack may earn a reputation for untrustworthiness and be punished in future elections. According to the logic of Fearon (1994), leaders' need to maintain consistency between their words and deeds should lend credibility to their peace promises. Therefore, it is important to test if democracies have a superior signaling capability across the full spectrum of hostile to benign signals.

The second shortcoming of this literature is that it tends to understand the democratic advantage as something intrinsic to democratic institutions. But the independent role of prior beliefs about a signaler's inherent aggressiveness may be a significant factor. Individuals have their own belief systems or knowledge structures about other states, including their regime types, alliances with their own country, etc. Some people, for example, having grown up in a democratic environment, may share faith in this political system, foster a sense of in-group identity with other democracies, and might endorse the notion that democracies do not fight each other (

Figure 1. Variations of credibility gap depending on prior beliefs



Wendt 1994; Risse-Kappen 1995; Alexander, Levin, and Henry 2005; Lacina and Lee 2013) as some constructivists argue. Once primed by relevant information, such as the regime-type of target countries, and then requested to assess state intentions, prior beliefs associated with some regime types may affect the assessment of state intentions and subsequently the credibility of state signals. This suggests that perceptions of a state's credibility may be shaped by prior beliefs about its regime type in addition to inferences based on its own behavior.

[Equation 1]

$$\underbrace{ND_{sig} - D_{sig}}_{\text{Credibility Gap (Democratic Credibility)}} = \underbrace{\{(D_{NoSig} - D_{sig}) - (ND_{NoSig} - ND_{sig})\}}_{\text{Perceived Signaling Capability Gap}} + \underbrace{\{ND_{NoSig} - D_{NoSig}\}}_{\text{Prior Belief Difference}}$$

*Note: ND: Non-democracy, D: Democracy. 'sig' represents signals, including both benign (ben) and hostile (hos) signals. 'NoSig' indicates no signal.

Presented in Equation 1 (and illustrated in Figure 1) is a new framework showing the credibility gap between two signalers that consists of the perceived signaling capability gap (as assumed by democratic credibility theory) but now includes benign and hostile signals as well as the effect of prior beliefs. Equation 1 illustrates the composition of this democratic advantage in signaling. Each term in Equation 1 indicates the level of aggressiveness assessed by observers when a signaler sends a specific type of signal. ND and D in Equation 1 represent the regime type of signalers: non-democratic and democratic, respectively, indicating that observers of state signals receive information and thereby know the regime types of the signalers. The subscript 'sig' represents signals, which include both benign (denoted as 'ben') and hostile signals (denoted as 'hos'), and the strength of signals is assumed to be equal regardless of who sends the signal. For example, 'ND_{hos}' denotes the level of perceived aggressiveness when a non-democratic signaler issues a hostile signal, whereas 'D_{hos}' indicates the perceived aggressiveness when a democratic signaler issues a hostile signal with the same strength. Consequently, 'ND_{hos} - D_{hos}' measures the difference in perceived aggressive intentions when both democracies and non-democracies send

identical hostile signals. For signals to be considered credible, hostile signals should be assessed as aggressive and benign signals as less aggressive. Thus, if ' $ND_{hos} - D_{hos} < 0$ ', it suggests that observers perceive democracies as more aggressive, and the larger this difference, the more they trust signals from democracies. Similarly, ' $ND_{ben} - D_{ben}$ ' measures the difference in perceived aggressive intentions when both democracies and non-democracies send identical benign signals. If ' $ND_{ben} - D_{ben} > 0$,' it indicates that observers perceive democracies as less aggressive, and the larger this difference, the more the observers trust signals from democracies.

The credibility gap consists primarily of two elements: the perceived signaling capability gap between state actors and the difference in *prior beliefs* about inherent aggressiveness of the actors. The first element affects democratic credibility by influencing the slopes of lines, while beliefs in the second element affect democratic credibility by changing the Y-intercept of lines, as shown in Figure 1. Specifically, the perceived signaling capability gap represents the difference in how observers evaluate the signaling capabilities of democracies versus non-democracies. The assessment of a signaler's signaling capability depends on both the signaler's actual capability and observers' beliefs about it. Together, these factors affect the slopes in Figure 1. This is measured by comparing assessments of a signaler's intentions when state actions (i.e., signals) are observed versus when they are not. Thus, a democracy's signaling capability is calculated by subtracting the assessments of a democracy's aggressive intentions when a signal is sent (D_{sig}) from the assessments when no such signal is sent (D_{NoSig}), resulting in $D_{NoSig} - D_{sig}$ in Equation 1. Likewise, $ND_{NoSig} - ND_{sig}$ represents the degree to which a non-democracy's signals update observers' prior beliefs about the non-democracy's intentions. When the signals from two signalers are identical in content and strength, the disparity in the credibility of their signals, represented as $(D_{NoSig} - D_{sig}) - (ND_{NoSig} - ND_{sig})$, equates to observers' assessments of the gap in *signaling capabilities* between the two signalers.

Differences in beliefs about the inherent aggressiveness of regime types can be measured based on the assessment of aggressiveness with no information except the regime type ($ND_{NoSig} - D_{NoSig}$). When observers are requested to infer signalers' intentions given the regime-type information, their beliefs about the signalers' aggressive or peaceful nature will be primed by their belief systems and may respond differently to different regime types depending on their beliefs. This prior belief difference, if any, increases or decreases the overall credibility gap of state signals

($ND_{sig} - D_{sig}$) by moving the Y-intercept of lines in Figure 1. In fact, the impact of pre-existing beliefs on information processing is a well-established phenomenon recognized in both cognitive psychology and political science (See, for instance, Festinger 2001; Zaller 1992; Jervis 2006; Yarhi-Milo 2013). For instance, the favorable attitudes of Tony Blair and Bill Clinton towards democracies, as opposed to non-democracies, led their countries to be more cooperative with democratic countries. However, due to Blair's greater hostility towards non-democracies compared to Clinton, the UK's foreign policy exhibited greater aggression towards non-democracies than the US (Schafer and Walker 2006).

Figure 1 presents three graphs extending democratic credibility theory to include benign signals. It illustrates how the observability of democratic advantage in signaling can vary depending on observers' prior beliefs when one signaler (Signaler 1) is evaluated as having superior signaling capability over another (Signaler 2). In Figure 1, the X-axis represents different signal conditions wherein observers encounter various types of signals: benign, no signal, and hostile. The "Benign" and "Hostile" conditions embody the types of intentions encapsulated in state signals, whereas "No Signal" refers to the condition in which observers do not receive any additional information about a signaler's intentions beyond their regime type. The Y-axis displays the perceived aggressiveness of the signaler by the observers. The blue dotted line represents the signals from Signaler 1 (e.g., a democracy), whereas the solid black line denotes the signals from Signaler 2 (e.g., a non-democracy).²

Figure 1(a) presents the basic assumption of democratic credibility theory in which there are no differential prior beliefs between two signalers' aggressiveness ($ND_{NoSig} - D_{NoSig} = 0$). Both signalers are assessed as equally aggressive (γ), as shown in the "No Signal" condition. Signaler 1 (e.g., democracy) has a steeper slope, indicating a perceived superior signaling capability. Thus, Signaler 1's benign signals are interpreted as more benign, and their hostile signals as more aggressive, than those from Signaler 2. In Figure 1(a), where an observer receives benign signals of equal intensity from both signalers in the 'Benign' condition, observers gauge Signaler 1

² Changes in the y-variable depending on the change of signal conditions on the x-axis are described as a continuous line to highlight the shifts of the observers' evaluation even though signal conditions are categorical variables.

(democracy)’s aggression level as D_{ben} while it estimates Signaler 2 (non-democracy)’s aggression level as ND_{ben} ($D_{ben} < ND_{ben}$). The benign signals from Signaler 1 are assessed as more credible than comparable signals from Signaler 2 by a difference of $ND_{ben} - D_{ben}$. This credibility gap is both conceptually and mathematically equivalent to the perceived signaling capability gap between the two signalers ($= (\gamma - D_{ben}) - (\gamma - ND_{ben})$). Similarly, in the ‘Hostile’ condition of Figure 1(a), where an observer receives hostile signals from both signalers, the credibility gap, $D_{hos} - ND_{hos}$, aligns with the perceived difference in the two signalers’ signaling capabilities ($= (D_{hos} - \gamma) - (ND_{hos} - \gamma)$).

Suppose Signaler 1 has superior signaling capability—and is perceived as such—compared to Signaler 2, as illustrated in Figure 1(a), but now the observers hold distinct prior beliefs about the nature of the two signalers. An observer’s prior beliefs pertain to their views about a signaler’s intentions before receiving any information about that signaler. The difference in prior beliefs about the two signalers’ aggressiveness is quantified as the difference in the observer’s inference about the intentions of the two signalers in the ‘No Signal’ condition: $ND_{NoSig} - D_{NoSig}$. Figure 1(b) illustrates a scenario where an observer holds a relatively positive prior belief regarding Signaler 1 ($ND_{NoSig} - D_{NoSig} > 0$), whereas Figure 1(c) portrays a scenario where an observer has a relatively negative prior belief regarding Signaler 1 ($ND_{NoSig} - D_{NoSig} < 0$). As demonstrated in Figures 1(b) and 1(c), prior beliefs induce a distinct dynamic in the credibility of different types of signals.

Figure 1(b) illustrates the case where Signaler 1 has a more substantial advantage in conveying its *benign* intentions (compare the size of $ND_{ben} - D_{ben}$ in Figures 1(a) and 1(b)). However, this advantage does not extend to Signaler 1 signaling its *hostile* intentions (compare the size of $D_{hos} - ND_{hos}$ in Figures 1(a) and 1(b)). For benign signals, the credibility gap is large because the difference in prior beliefs about the two signalers’ aggressiveness is positive ($ND_{NoSig} - D_{NoSig} > 0$) and Signaler 1 is assessed to have superior signaling capability ($(D_{NoSig} - D_{ben}) - (ND_{NoSig} - ND_{ben}) > 0$). Conversely, for hostile signals, the overall credibility advantage of Signaler 1 ($ND_{hos} - D_{hos}$) should weaken or vanish because the difference in prior beliefs ($ND_{NoSig} - D_{NoSig} > 0$) offsets Signaler 1’s superior signaling capability ($(D_{NoSig} - D_{hos}) - (ND_{NoSig} - ND_{hos}) < 0$). Therefore, Figure 1(b) indicates that when an observer holds a more positive prior belief towards Signaler 1 than towards Signaler 2, Signaler 1 will have less or no apparent advantage in signaling its aggressive intentions compared to when an observer does not hold such a positive prior belief.

Hence, even if an observer evaluates that Signaler 1's signaling capability surpasses that of Signaler 2, they may perceive no significant difference in the overall credibility of hostile signals from the two signalers.

Figure 1(c) presents the opposite case where an observer harbors a somewhat negative prior belief towards Signaler 1 ($ND_{NoSig} - D_{NoSig} < 0$). In this scenario, Signaler 1 should demonstrate a larger advantage in conveying *hostile* intentions (note the difference in $ND_{hos} - D_{hos}$ between Figure 1(a) and 1(c)). In line with Equation 1, a considerable credibility gap ($ND_{hos} - D_{hos}$) is expected for hostile signals. This gap emerges as the observer views Signaler 1 (e.g., a democracy) as more aggressive ($ND_{NoSig} - D_{NoSig} < 0$) and credits it with superior signaling capability ($(D_{NoSig} - D_{hos}) - (ND_{NoSig} - ND_{hos}) < 0$). Given the observer's preconceived belief of the signaler's aggressiveness, Signaler 1's hostile signals should be deemed more credible.

However, this advantage may diminish when signaling *benign* intentions (compare $ND_{ben} - D_{ben}$ in Figures 1(a) and 1(c)). Signaler 1's credibility advantage should wane or disappear altogether for benign signals ($ND_{ben} - D_{ben}$) when the observer harbors a negative prior belief towards Signaler 1 (refer to Figure 1(c)). This negative prior belief difference ($ND_{NoSig} - D_{NoSig} < 0$) might counteract the assessment of Signaler 1's superior signaling capability ($(D_{NoSig} - D_{ben}) - (ND_{NoSig} - ND_{ben}) > 0$). As a result, despite acknowledging Signaler 1's superior signaling capability, the observer may perceive no significant difference in the assessment of intentions between the benign signals of the two signalers.

To sum up, Figures 1(a) to (c) demonstrate how prior beliefs about a signaler's peaceful or aggressive nature can systematically impact the credibility of signals. This new model helps bridge the debate between scholars who have found evidence for democratic credibility theory and those who have not by suggesting when and why a democracy's resolve for war may be unobservable. The following section outlines hypotheses derived from this theory.³

³ Refer to Online Appendix A for a summary of the main arguments and predicted outcomes from diverse perspectives on democratic credibility.

Hypotheses

This study tests the presence of democratic credibility by investigating public opinions in the United States and China. Many observers suggest these two states—which are respectively the most powerful democracy and non-democracy in the world—are locked in a cycle of military competition in East Asia with huge potential for crisis instability. Thus, it is important to better understand the crisis signaling between the most powerful democratic and non-democratic states (e.g., Kertzer, Brutger, and Quek 2024). Despite their difference in regime type, public opinion plays a significant role in shaping foreign policy in both countries. In the United States, citizens influence their government's foreign policy preferences through politicians' concerns about election outcomes (Yarhi-Milo, Kertzer, and Renshon 2018; Tomz, Weeks, and Yarhi-Milo 2020). Additionally, despite the common belief that public opinion holds little importance in authoritarian regimes, contemporary researchers are shedding light on how public sentiment can influence foreign policies in the regimes (For its specific importance in the context of China, see, for example, (Reilly 2012; Weiss 2013; Li and Chen 2021). Therefore, understanding how the public in both countries perceives diplomatic statements and actions of other states is crucial.

Five sets of hypotheses were created to test the theory proposed in this study. The first set of hypotheses focuses on a key aspect of democratic credibility, as proposed by early signaling theorists in IR (Fearon 1994; Schultz 1998). These hypotheses examine whether democracies are perceived to have a superior signaling capability, technically expressed as whether $(D_{\text{NoSig}} - D_{\text{sig}}) - (ND_{\text{NoSig}} - ND_{\text{sig}})$ in Equation 1. As noted earlier in this paper, this notion has often been challenged by some scholars in IR (Weeks 2008; Trager 2010; Snyder and Borghard 2011; Trachtenberg 2012; Downes and Sechser 2012). The study investigates these hypotheses both when the observers are American and when they are Chinese.

Hypothesis 1-1a. Holding other factors constant, Americans are more likely to assess the signaling capabilities of a democracy as superior to those of a nondemocracy.

Hypothesis 1-1b. Holding other factors constant, Americans are *not* likely to assess the signaling capabilities of a democracy as superior to those of a nondemocracy.

Hypothesis 1-2a. Holding other factors constant, Chinese are more likely to assess the signaling capabilities of a democracy as superior to those of a nondemocracy.

Hypothesis 1-2b. Holding other factors constant, Chinese are *not* likely to assess the signaling capabilities of a democracy as superior to those of a nondemocracy.

The second set of hypotheses aims to investigate the critical role of prior beliefs about the nature of signalers' aggressiveness ($ND_{NoSig} - D_{NoSig}$, as detailed in Equation 1 in the previous section). US citizens are likely to hold positive prior beliefs about democracies, having been educated throughout their upbringing about the merits of the democratic political system. The democratic peace literature reflects the favorable beliefs or attitudes that citizens of democratic nations have towards other democratic states (Doyle 1986; Maoz and Russett 1993; Risse-Kappen 1995). In contrast, the beliefs held by Chinese citizens towards democracies are less predictable and likely vary. However, a recent poll conducted among the Chinese population in 2022 suggests their negative beliefs and attitudes towards democracies like the United States, India, and Japan, in contrast to their favorable attitudes towards nondemocracies such as Russia, Pakistan, and North Korea (Turcsanyi et al. 2022). The following set of hypotheses is derived from these observations.

Hypothesis 2-1a. Holding other factors constant, Americans are likely to perceive democracies as more peaceful than non-democracies.

Hypothesis 2-1b. Holding other factors constant, Americans are *not* likely to perceive democracies as more peaceful than non-democracies.

Hypothesis 2-2a. Holding other factors constant, Chinese are likely to perceive democracies as more aggressive than non-democracies.

Hypothesis 2-2b. Holding other factors constant, Chinese are *not* likely to perceive democracies as more aggressive than non-democracies.

Consider how Americans might assess the credibility of *benign* signals from a democracy versus those from a non-democracy ($ND_{ben} - D_{ben}$ in Equation 1). Coupled with their assessment of a democracy's superior signaling capability (due to mechanisms such as democratic leaders facing greater audience costs etc.) and positive prior beliefs about democracies (as inherently more peaceful), Americans may find a democracy's benign signals more credible than those from a non-democracy (see the credibility gap in the "Benign" condition of Figure 1(b)). The theory, however, anticipates different results for *hostile* signals ($ND_{hos} - D_{hos}$ in Equation 1). Although Americans may recognize the superior signaling capability of a democratic signaler, their belief that democracies tend to be peaceful may compromise the credibility of the democratic signaler's hostile signals (see the credibility gap in the "Hostile" condition of Figure 1(b)). For example, Trump's outlandish threats to North Korea triggered intense invasion fears, while his criticism of Canada—calling it a tyranny and even bringing up the destruction of the White House in the War of 1812 to justify tariffs—did not trigger any significant invasion fears in Canada (Acosta and Paula 2018). The following third set of hypotheses has been formulated based on this insight.

Hypothesis 3a: Americans are likely to assess a democracy's benign signals as more credible than those from a non-democracy, but unlikely to assess a democracy's hostile signals as more credible than those from a non-democracy.

Hypothesis 3b: Americans are likely to perceive no significant difference in the credibility of benign or hostile signals between democracies and non-democracies.

Predicting the responses of Chinese observers is more complex due to ambiguities in their perceptions of democratic signaling capabilities and their beliefs about the aggressiveness of democracies. Several scenarios are possible. In one, Chinese observers may perceive democracies

as having superior signaling capabilities compared to non-democracies and hold significantly negative prior beliefs about them, corresponding to Figure 1(c). In the opposite scenario, Chinese observers neither recognize the superior signaling capabilities of democracies nor hold any particular prior beliefs about their aggressiveness. Other scenarios would fall somewhere between these two extremes. Two hypotheses have been derived from these scenarios.

Hypothesis 4a: Chinese are likely to assess a democracy's *hostile* signals as more credible than those from a non-democracy, but unlikely to assess a democracy's *benign* signals as more credible than those from a non-democracy.

Hypothesis 4b. Chinese are likely to perceive no significant difference in the credibility of benign or hostile signals between democracies and non-democracies.

The final hypothesis concerns the mode of signaling. Although the theory proposed in this study does not distinguish between them, it is empirically important to determine whether words or actions constitute a more effective mode of signaling and consequently affect democratic credibility. As expected by the literature, a democratic leader's threatening words may be perceived as more credible (Fearon 1994), while signs of autocracy's military actions might be taken as seriously as those of a democracy (Slantchev 2005).

Hypothesis 5a: A democratic advantage in signaling will be more apparent when diplomatic statements, rather than military actions, are used by signalers to convey their intentions.

Hypothesis 5b: There is no significant difference in the effectiveness of diplomatic statements versus military actions used by signalers, regardless of their regime type.

Experimental Design

A survey-based experiment was employed to test the proposed theory. This methodology has been extensively discussed by scholars using similar approaches, so its pros and cons will not be repeated here (For the utility of survey experiments, see Tomz 2007; Tomz and Weeks 2013; Yarhi-Milo, Kertzer, and Renshon 2018; Kertzer, Renshon, and Yarhi-Milo 2021).

This study shares common concerns associated with its methodology, namely external validity and generalizability. Several points potentially mitigate these concerns. First, an increasing number of studies in IR, particularly those focusing on signaling perception, have found no significant discrepancies between the responses of political elites and the general population (Yarhi-Milo, Kertzer, and Renshon 2018; Kertzer, Renshon, and Yarhi-Milo 2021; Kertzer and Renshon 2022; Renshon, Yarhi-Milo, and Kertzer 2023).⁴ Furthermore, the inclusion of Chinese Communist Party members in the Chinese sample enables the inference of Chinese elites' responses from the subgroup. In China, only a fraction of selected participants can be members of the Communist Party (only 1 in 16 Chinese). Once accepted as a party member, they “hold nearly every top position in government, military, education, state-owned enterprises, health care, and banking” (McMorrow 2015).

Second, the original study was initially fielded in the United States and then, to enhance the generalizability of the proposed theory and comparison with the experimental results of the US sample, it was expanded to include citizens in China.⁵ First, the study was conducted with a sample of 384 American adults in English, recruited via Amazon Mechanical Turk in February 2015 (regarding the representativeness of the sample recruited by MTurk, see, for example, Berinsky, Huber, and Lenz 2012 and Kertzer, Renshon, and Yarhi-Milo 2021's Online Appendix).⁶ Then, a

⁴ Most notably, in Kertzer (2022)'s meta-analysis of paired experiments involving political elites and public samples, only 19 (11.7%) out of 162 treatment effects from 48 paired experiments he compared showed significant differences in magnitude between the two groups, and only 3 (1.9%) differed in sign.

⁵ See, for example, Levin and Trager (2019) and Feng, He, and Li (2019) for how the Chinese perspective contributes to generalizability of the theory.

⁶ The size of the sample was chosen to detect 0.3-sized effects of the experimental manipulation with a power of 0.8.

parallel experiment was conducted in 2022 in Chinese with a sample of 1,740 Chinese adults aged 18 years or older using Qualtrics, a major survey company.⁷ The company employed sampling strategies in line with population targets from the 2021 Chinese national census, matching for age and gender. The experiment with the Chinese participants mirrored the one with the American participants, with minor adjustments made to maintain scenario plausibility within the Chinese context.

The experiment commenced with a typical introductory paragraph, (Tomz 2007; Tomz and Weeks 2013), which was modified for this study as follows: “The country described below is hypothetical and does not exist. That said, the scenario described here is one the United States [China] has faced and may have to deal with again in the future.” Participants were then presented with a scenario, which started with “The government of the United States [China] needs to take a position on a territorial dispute between Country X and a US ally [China’s important partner country] in the region. The ally [partner country] is important to US [China’s] security, has strong diplomatic ties with the US [China], and is a major trading partner. The central aims of US [China’s] foreign policy in this region are to protect the US ally [China’s partner country] and to prevent the emergence of a hostile regional power that can threaten US [China’s] security.” To eliminate potential bias, the hypothetical country was referred to as “Country X” rather than a specific country name.⁸ Additionally, the dispute was framed as being between a US *ally* (China’s partner country) and Country X, rather than directly involving the US [China], to avoid associating the hypothetical country with any real-world adversaries of the United States [China] (See Online Appendix B2 for the complete scenario).

Participants were next introduced to a scenario in which Country X and the US ally (China’s partner country) were historically regional rivals. The US (Chinese) government is uncertain of Country X’s intentions, such as whether Country X is a peaceful country that is not interested in occupying the territory of the US ally (China’s partner country) or an aggressive

⁷ A larger number of the Chinese were recruited than Americans to obtain a comparable size of a political elite sample in China.

⁸ For justification of using a hypothetical scenario and an unnamed country instead of real situations and a specific country, see, for example, Brutger et al. 2023.

country that wants to use the border as a launching pad for further territorial expansion. Participants then received three separate treatments in signaler's regime type and signal types: the regime type (democracy or non-democracy), the intentions embedded in the state signals (peaceful or aggressive), and the mode of signals (diplomatic statements or military actions). These treatments were independently and randomly varied. The combination of two types of intentions and two signaling modes resulted in four signal types. Including a control condition, a total of five signal conditions were created. Consequently, the experiment employed a 2 (regime type) x 5 (signal conditions) between-subjects design. The control condition served as a reference point against which the treatment effect of signals could be measured on the perception of the signaler's intentions.

First, *Regime Type* indicates the regime type of the signaler: either democracy or non-democracy. For American participants, democracy was described as regularly holding “free and fair elections,” while a nondemocracy was described as a regime that does not hold “free and fair elections.” For Chinese participants, Country X was simply described as either “a democracy and shows every sign that it will remain a democracy” or “not a democracy and shows no sign of becoming a democracy” following previous studies (Tomz and Weeks 2013; Bell and Quek 2018). This difference in wording was necessary to avoid provoking respondents by providing a definition that openly challenged the widely held Chinese perception of democracy, which deemphasizes the role of “political contestation” (Hu 2018). However, despite the difference in wording between the two surveys, it is likely that Chinese respondents' ideas about democracy are comparable with those of the US respondents. As noted by Hu (2018, 1441), “Chinese understanding of democracy has reached to a certain degree of consensus that is closer to universally-shared idea of democracy rather than being culturally distinctive.”⁹ Second, *Intentions* specifies a signaler's intentions embedded in benign or hostile state signals. Benign signals or gestures of conciliation are intended to convey peaceful intentions of the signaler. Conversely, hostile signals or gestures of a

⁹ Moreover, to address the concern that differences in the Chinese individual understanding of democracy, instead of the treatments, might affect their perceptions and attitudes in the experiment, a comparison was made between subgroups that identify China as a democracy and those that do not, but no significant difference was found.

willingness to fight are intended to broadcast aggressive intentions. Third, *Signaling Mode* encompasses diplomatic statements and military actions.

The combination of “*Intentions*” and “*Signaling Mode*” resulted in four distinct types of signals: benign-word, benign-military action, hostile-word, and hostile-military action. Along with the no-signal condition, there were five signal conditions in total. Participants in Group 1 (benign-military action) received a description of a country that “has withdrawn most (about 75%) of its army that had been deployed near the border territory” and “substantially decreased the weapons that had been deployed in the same area by about 75%, removing most of its tanks, attack helicopters, and missiles.” Those in Group 2 (benign-word) were presented with the statement, “The government of Country X has often pledged to peacefully resolve the dispute with the US ally.” Participants in Group 3 (control) did not receive any information about the signaler’s intentions. Those in Group 4 (hostile-word) encountered the statement, “The government of Country X has often threatened to use force in the dispute with the US ally.” Participants in Group 5 (hostile-military action) read that Country X “has increased its army deployment by a huge amount (about 75%) near the border territory” and “substantially increased the weapons that had been deployed in the same area by about 75%, deploying a large number of tanks, attack helicopters, and missiles.”

After being presented with information about the hypothetical state, participants were asked two attention-check questions.¹⁰ They were then requested to infer the intentions of the hypothetical state, evaluate the threat posed by the state to their own country’s security and interests, and express their support level for deploying military force against the state in the scenario. Inferences about state intentions served as the primary outcome measure, with threat assessments and foreign policy preferences as secondary outcome measures. Measuring observers’ inferences about state intentions offers a more direct method of evaluating their assessments of the credibility of state signals than secondary measures, which may be influenced by factors including relative power and observers’ individual dispositions such as attitudes toward war. However,

¹⁰ Recent studies highlight the critical role of participant attention within experimental settings, showing that inattentive participants undermine the quality of data from survey experiments (Ternovski et al. 2022; Peyton, Huber, and Coppock 2022).

observers are more inclined to fear and support the use of force against a state once they perceive it as more aggressive and hostile, which justifies the use of secondary measures to assess the credibility of state signals. These three dependent variables (DVs) were measured on a five-point Likert scale. The lowest score, 1, represented very peaceful intentions, no perceived threat, and strong opposition to force deployment. Conversely, the highest score, 5, signified very aggressive intentions, an extremely high threat level, and strong support for deploying troops. Finally, demographic and dispositional questions were asked, covering respondents' political knowledge, gender, age, education, ideology, and income, among other factors.

Results

US Sample

The experiment examined whether American participants place more trust in democratic signals than non-democratic ones. It focused on three primary measures: (1) the presence of any discernible democratic advantage or disadvantage in signaling, (2) the assessment of the gap in signaling capabilities between a democracy and a non-democracy, and (3) the disparity in prior beliefs about aggressiveness that Americans hold about the two regime types. The results are based on an analysis of responses from those who successfully passed the attention check (N=341). These results are not significantly different from the outcomes when including the results from those who failed the test.¹¹

Figure 2 presents the average responses of participants (group means) to various types of signals—benign (left), hostile (right), or no-signal (middle)—from a democracy and a non-democracy and their respective 95% confidence intervals. Table 1 offers a detailed comparison of how American participants perceive the performances of democracies and non-democracies in crisis communication, while showing how the assessment of credibility gaps between signalers are divided. The table examines any existing credibility gaps, which consist of the perceived signaling capability gap and prior belief difference between the two types of governments.

¹¹ Due to the 10-page limit on appendix material, relevant graphs are omitted but available upon request.

For the purposes of this analysis, Y_{ijk} is used to denote the response from individual i , under a specific signal treatment condition $j = \{0, 1, 2\}$, where 0, 1, and 2 represent information about signal types such as benign, no signal, and hostile signals, respectively. Additionally, $k = \{0, 1\}$ indicates the information about the regime type of signalers, with 0 representing a non-democracy and 1 representing a democracy. Thus, the treatment effect of benign signals among American observers is the difference between the treatment (benign) and the control condition (no signal), measured as $E[Y_{i11} - Y_{i01}]$ for a democracy and $E[Y_{i10} - Y_{i00}]$ for a non-democracy. Consequently, the average gap in the assessment of signaling capabilities for benign signals between the two regimes is the difference-in-difference ($E[Y_{i11} - Y_{i01}] - E[Y_{i10} - Y_{i00}]$). Columns (B), (C), and (D) in Table 1 present these measurements in order. Similarly, the average prior belief difference regarding the aggressive nature of democracies and non-democracies is measured as $E[Y_{i10} - Y_{i11}]$ and is recorded in columns (E) and (J).

The overall advantage or disadvantage that democracies have in signaling their benign intentions is measured as $E[Y_{i00} - Y_{i01}]$ and is recorded in column (A). Likewise, the average assessment of a signaler's signaling capability for hostile signals is measured as $E[Y_{i21} - Y_{i11}]$ for a democracy and $E[Y_{i20} - Y_{i10}]$ for a non-democracy. The average gap in the assessment of signaling capabilities for hostile signals between the two regimes is thus $E[Y_{i21} - Y_{i11}] - E[Y_{i20} - Y_{i10}]$. Columns (G), (H), and (I) represent these in order. The overall advantage or disadvantage that democracies have in signaling their aggressive intentions is measured as $E[Y_{i20} - Y_{i21}]$ and is recorded in column (F).

Figure 2. Americans' Assessments of Benign and Hostile Signals

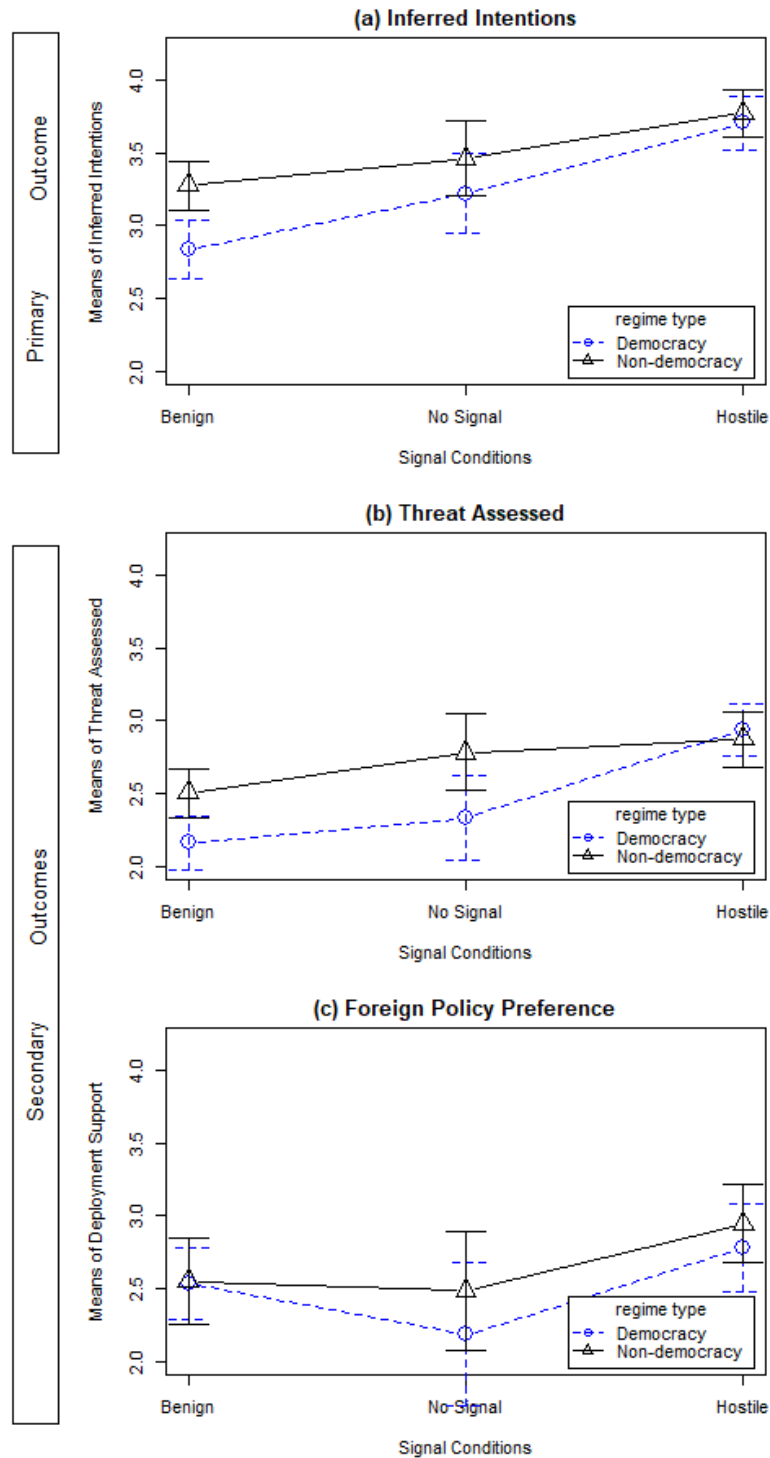


Table 1. Democratic Advantage and Its Components when Americans are Observers

Signal Types	DV	Democratic Credibility (Credibility Gap)	Perceived D's SC	Perceived ND's SC-	Perceived SC Gap	Prior Belief Difference
		$ND_{ben} - D_{ben}$	$D_{NoSig} - D_{ben}$	$ND_{NoSig} - ND_{ben}$	(B)-(C)	$ND_{NoSig} - D_{NoSig}$
		(A)	(B)	(C)	(D)	(E)
Benign Signals	Inference	.440***	.387*	.184	.203	.237
	Threat	.343**	.169	.277	-.108	.450*
	FPP	.016	-.349	-.064	-.285	.301
		$ND_{hos} - D_{bos}$	$D_{NoSig} - D_{hos}$	$ND_{NoSig} - ND_{hos}$	(G)-(H)	$ND_{NoSig} - D_{NoSig}$
		(F)	(G)	(H)	(I)	(J)
Hostile Signals	Inference	.071	-.481**	-.315*	-.166	.237
	Threat	-.064	-.604***	-.089	-.515**	.450*
	FPP	.162	-.596*	-.457	-.139	.301

Note 1: *: $p < .05$, **: $p < .01$, ***: $p < .001$.

Note 2: ND: Nondemocracy, D: Democracy, SC: Signaling Capability, Subscript ben: benign signals, hos: hostile signals, NoSig: No Signal

There are several key findings. First, consistent with Hypothesis 1-1a, American participants were inclined to assess the signaling capabilities of a democracy as superior to those of a non-democracy, particularly when confronted with hostile signals, as early signaling theorists such as Fearon (1994) and Schultz (1998) postulated (compare columns (B) and (C), and (G) and (H) in Table 1). A democracy's hostile signals were perceived as credible, compared to when there were no such signals, in intention inference ($p < .01$), threat assessment ($p < .001$), and foreign policy preference ($p < .05$). In contrast, a non-democracy's signals failed to cause American participants to significantly update their initial beliefs about a non-democracy's intentions, unless tasked with inferring its hostile signals (see columns (C) and (H)). The differences in the assessment of signaling capabilities of the two regimes (see columns (D) and (I)), however, were statistically significant only in threat assessments of hostile signals ($p < .01$).

Second, there is consistent evidence supporting Hypothesis 2-1a that Americans are likely to perceive a democracy as more peaceful than a non-democracy, given no information about intentions except the regime type of the signaler. As shown in the middle of the three graphs in Figure 2, No-Signal condition, participants consistently perceived a democracy as more peaceful and less threatening than a non-democracy without further information. However, the differences were statistically significant only when assessing threats (see columns (E) or (J) in Table 1).

Most interestingly, consistent with Hypothesis 3a, American participants, on average, tended to trust a democracy's *benign* gestures more than those of a non-democracy (as shown in two out of the three DVs), but they did not tend to take a democracy's *hostile* gestures more seriously than those of a non-democracy. American participants assessed a democracy's benign signals as significantly more benign ($p < .001$) and less threatening ($p < .01$) than those of a non-democracy when they received information about the democracy's peaceful intentions, as shown in the 'benign' conditions on the left side of Figures 2(a) and (b) (see also column (A) of Table 1). In contrast, when participants were asked to evaluate hostile signals, there was no significant difference between the regime types, as shown in the 'hostile' conditions on the right side of Figures 2(a), (b), and (c) (see also column (F) in Table 1).

These overall results partially support the democratic credibility thesis in that democracies have the advantage of communicating their intentions. However, this advantage is limited to

instances when democracies want to transmit *benign*, not hostile signals. What explains this result? Table 1 reveals that the democratic advantage in signaling is the result of a combination of differences in the assessment of signaling capabilities and prior belief differences between the two regimes. For benign signals, a democratic advantage was discernible in two of the three DVs (see column A in Table 1). In the case of intention inference, for example, a democracy's superior signaling capability (.203) and Americans' favorable preconceptions toward a democracy (.237) both contributed approximately equally to the overall democratic advantage (.440). Conversely, no clear democratic advantage emerged under 'hostile' conditions (see column (F)). American participants' positive prior beliefs toward a democracy were strong enough to almost negate their perception of a democracy's superior signaling capability. This result partially explains, at a micro-level, why it may be challenging to discern democratic credibility in large-N data or historical contexts, even when democracies indeed are perceived to be more credible than non-democracies. The favorable beliefs held by citizens of democracies towards other democracies may obscure any democratic advantage in signaling resolve or hostility.

It is noteworthy that no evidence of democratic credibility was observed when Americans were asked about their stance on force deployment. This indicates that even though Americans may view other democracies as more peaceful and less threatening, such perceptions may not necessarily influence their opinions on force deployment. This suggests individual respondent characteristics such as political ideology and political knowledge may play more pivotal roles in shaping foreign policy preferences than just the information about the regime type of a signaler.¹²

Furthermore, the analysis of the credibility of signals from a democracy and a non-democracy, depending on the signaling mode, reveals that a democracy's advantage over a non-democracy in expressing its intentions is more pronounced through *words* rather than *deeds* for benign signals (see Figure C in the Online Appendix). However, this distinction does not apply to hostile signals. This finding partially supports Hypothesis 5a. American participants on average

¹² A fuller analysis indicates that the explanatory power of independent variables in foreign policy preference diminishes by 64% (R-squared decreases from 0.11 to 0.04) when variables of political ideology and political knowledge are excluded from the regression equation.

trusted the peace promises made by a democratic government more than those made by a non-democratic government. Yet, this democratic advantage diminishes when the government communicates its message through strong military actions. This is primarily because strong military actions from a non-democratic government are deemed as credible as those from a democracy as Slantchev (2005) claimed.

Chinese Sample

The experiment conducted on Chinese citizens aimed to examine whether the democratic credibility observed among US citizens would be similarly found in this group. Figure 3 shows the average responses from 1,318 Chinese participants, all aged 18 or older, who successfully completed the attention check. The analysis from this subset did not significantly differ from the results of the full sample (N=1,540) including those who failed the attention check.¹³

First, Chinese participants exhibited varying degrees of trust towards benign and hostile types of signals, depending on the regime type of the signaling party. When dealing with a democracy, they were more inclined to trust its benign signals and update their initial beliefs in response to these signals, but not in response to *hostile* ones (see columns (B) and (G) in Table 2). In contrast, neither benign nor hostile signals from a non-democracy significantly updated the prior beliefs of Chinese participants in general (see columns (C) and (H) in Table 2). As a result, for benign signals, Chinese participants appeared to assess a democracy's signaling capability as superior to that of a non-democracy, even though the difference was statistically significant only in threat assessment (see Column (D) in Table 2). However, Chinese participants seemed to assess a non-democracy's hostile gestures as significantly more aggressive than those of a democracy (see column (I) in Table 2). These results support Hypothesis 1-2a only in part, suggesting that Chinese citizens may assess democracies as superior to non-democracies in their capability only under a few conditions. These patterns of responses were similarly found among Chinese political elites.¹⁴

¹³ Relevant graphs are available upon request.

¹⁴ Relevant graphs and tables are available upon request.

Second, consistent with Hypothesis 2-2a, Chinese participants tended to harbor prior beliefs that democracies are more aggressive compared to nondemocracies, even though these differences did not reach statistical significance at conventional levels. The negative prior beliefs of Chinese respondents towards democracies are more evident among elites). Members of the Chinese Communist Party in the sample exhibited stronger negative beliefs towards a democracy than toward a nondemocracy in terms of intention inference ($p=.08$), threat assessments ($p<.05$), and support for the deployment of force ($p<.01$).

Third, in partial agreement with Hypothesis 4a, Chinese participants were unlikely to assess a democracy's benign signals as more credible than the same signals from a non-democracy (see column (A) in Table 2). It is because the perceived superiority of benign signals from a democracy (e.g., .268 in threat assessment) was negated by Chinese observers' pre-existing beliefs towards a democracy (e.g., -.231 in threat assessment). In addition, the Chinese, on average, did not perceive a democracy's hostile signals as more credible than those of a non-democracy either (see column (F) in Table 2). This is mostly because the Chinese did not perceive a democracy's capability to signal their aggressive intentions as superior to that of a non-democracy (see column (G) in Table 2), even though Chinese participants held somewhat negative prior beliefs about a democracy's intentions. The same patterns were observed among Chinese elites. Finally, contrary to Hypothesis 5a, democratic credibility was not more apparent in a democracy's diplomatic statements among the Chinese participants. In fact, no democratic advantage or disadvantage was observed in either diplomatic statements or military actions.

Further analyses were conducted on the subgroups within the Chinese sample.¹⁵ The responses of the members of the Chinese Communist Party in the subsample ($N=405$) were analyzed. The results indicate that the responses of these Chinese elites fundamentally align with those of ordinary Chinese citizens. Another analysis was carried out to address the potential concern that Chinese respondents may have different perceptions and attitudes towards state signals depending on their understanding of democracy. For example, those identifying China as a democracy might exhibit more or less favorable attitudes toward democracies and understand their signals differently. To mitigate this concern, the Chinese sample was divided into two groups:

¹⁵ Relevant analyses are available upon request.

those who identify their own country as a democracy and those who perceive it as a non-democracy. However, no significant differences were observed based on their understandings of democracy. Additionally, the characteristics of individual Chinese participants, such as age, education, family income, and Communist Party membership, account for more variation in the DVs than either the regime type of signalers or the content of the messages.

Figure 3. Chinese Assessment of Benign and Hostile Signals

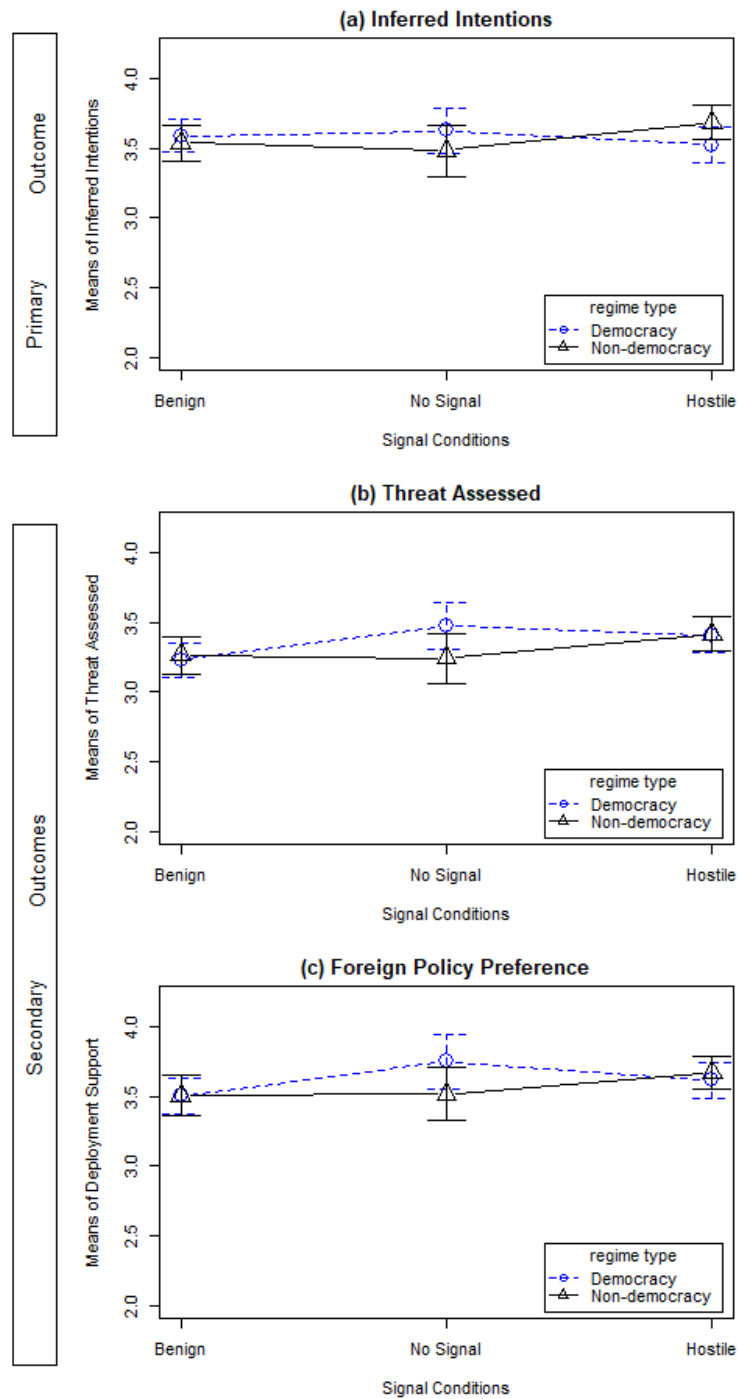


Table 2. Democratic Advantage and Its Components when the Chinese are Observers

Signal Types	DV	Democratic Credibility (Credibility Gap)	Perceived D's SC	Perceived ND's SC-	Perceived SC Gap	Prior Belief Difference
		$ND_{ben} - D_{ben}$	$D_{NoSig} - D_{ben}$	$ND_{NoSig} - ND_{ben}$	(B)-(C)	$ND_{NoSig} - D_{NoSig}$
		(A)	(B)	(C)	(D)	(E)
Benign Signals	Inference	-.052	.036	-.054	.090	-.142
	Threat	.037	.243*	-.025	.268*	-.231
	FPP	.002	.244*	.010	.234	-.232
		$ND_{hos} - D_{hos}$	$D_{NoSig} - D_{hos}$	$ND_{NoSig} - ND_{hos}$	(G)-(H)	$ND_{NoSig} - D_{NoSig}$
		(F)	(G)	(H)	(I)	(J)
Hostile Signals	Inference	.160	.102	-.200	.302*	-.142
	Threat	.007	.063	-.175	.238*	-.231
	FPP	.052	.131	-.153	.284	-.232

Note 1: *: $p < .05$, **: $p < .01$, ***: $p < .001$.

Note 2: ND: Nondemocracy, D: Democracy, SC: Signaling Capability, Subscript ben: benign signals, hos: hostile signals, No Sig: No Signal

Conclusion

Numerous scholars studying democratic credibility in signaling have struggled to produce supporting empirical evidence. Most critics of the democratic credibility theory believe that domestic audiences value policy outcomes more than the consistency of their leaders' words, or that non-democracies have the capability of making their signals as credible as democracies. Contrary to these critics, this paper presents a plausible theory explaining how the democratic advantage in signaling can exist but can remain unobserved. The findings of this study suggest that observers' prior beliefs about a signaler's aggressiveness, primed by regime type information, can systematically mask or amplify democratic credibility depending on the signal type.

Admittedly, there are some limitations to this study. First, even though it contributes to theoretically and empirically separating democratic credibility into its two component pieces—perception of signaling capabilities and prior beliefs—as a baseline study to understand the psychological dynamics behind the phenomenon, further studies about beliefs need to be conducted. For example, how do other kinds of beliefs interact with the assessment of signaling capabilities? Under which conditions do beliefs amplify or discount the perception of the signaling capabilities of democracies or non-democracies? Similarly, scholars in the future may also attempt to link observers' perceptions of democracies with the actual institutional features of democracy emphasized in classical accounts of democratic credibility. How do prior beliefs about democracy in general interact with specific institutional features? For example, is a signal from a coalitional government with strong oppositional parties (like many Western European governments) perceived as more credible than one from a democracy where one dominant party has maintained power for years and faces no serious opposition (like in Japan where the Liberal Democratic Party (LDP) dominated most of the post-war era)?¹⁶

This paper highlights the importance of micro-level dynamics in how observers' prior beliefs about signalers' regime type systematically influence the assessments of signal credibility and, in turn, democratic advantages in signaling. These findings bear significant implications for

¹⁶ For work in this direction, see Uzonyi, Souva, and Golder 2012.

several debates within IR. At the broadest level, this paper contributes theoretical insights to the debate between biased learning and Bayesian updating (e.g., Festinger 2001; Herrmann 2017; Kertzer, Rathbun, and Rathbun 2020) by providing evidence that supports both learning mechanisms. Observers were successfully able to engage in Bayesian updating of a signaler's intentions regardless of whether they held either positive or negative prior beliefs about these intentions. Second, the findings of this study suggest a plausible pathway that likely leads to democratic peace. Americans may be inclined to infer benign intentions from democracies *ex-ante*, and these prior beliefs may also help amplify whatever benign signals are made by fellow democracies. Similarly, Americans may discount the credibility of hostile signals from other democracies, potentially creating virtuous 'echo chambers' among democratic states where benign signals become louder and more persistent than hostile ones. Third, somewhat more pessimistically, this paper highlights how prior beliefs about regime types could act as a potential blind spot when policymakers are seeking to signal reassurance or resolve in a crisis. For example, Chinese citizens' and elites' negative beliefs about democracies' aggressiveness could cause them to discount benign signals from the United States, while believing that their own signals of resolve strike the right tone of assertiveness without being too assertive. On the other hand, US policymakers—believing that democracies are intrinsically peaceful—may overestimate the credibility of their own signals of reassurance (e.g., Kim 2022) while similarly exaggerating the aggressiveness of Chinese signals of resolve. This may lead to situations where benign signals must be costlier than necessary to be credible, or, tragically, to situations where signals of resolve trigger the very conflict they are seeking to deter.

References

- Acosta, Jim, and Newton Paula. 2018. "Exclusive: Trump Invokes War of 1812 in Testy Call with Trudeau over Tariffs." *CNN*, June 6, 2018.
<https://edition.cnn.com/2018/06/06/politics/war-of-1812-donald-trump-justin-trudeau-tariff/index.html>.
- Alexander, Michele G., Shana Levin, and P. J. Henry. 2005. "Image Theory, Social Identity, and Social Dominance: Structural Characteristics and Individual Motives Underlying International Images." *Political Psychology* 26 (1): 27–45.
- Bell, Mark S., and Kai Quek. 2018. "Authoritarian Public Opinion and the Democratic Peace." *International Organization* 72 (1): 227–42.
- Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.Com's Mechanical Turk." *Political Analysis* 20 (3): 351–68.
- Brown, Jonathan N., and Anthony S. Marcum. 2011. "Avoiding Audience Costs: Domestic Political Accountability and Concessions in Crisis Diplomacy." *Security Studies* 20 (2): 141–70.
- Brutger, Ryan, Joshua D. Kertzer, Jonathan Renshon, Dustin Tingley, and Chagai M. Weiss. 2023. "Abstraction and Detail in Experimental Design." *American Journal of Political Science* 67 (4): 979–95.
- Downes, Alexander B., and Todd S. Sechser. 2012. "The Illusion of Democratic Credibility." *International Organization* 66 (3): 457–89.
- Doyle, Michael W. 1986. "Liberalism and World Politics." *American Political Science Review* 80 (4): 1151–69.

- Fang, Songying. 2008. "The Informational Role of International Institutions and Domestic Politics." *American Journal of Political Science* 52 (2): 304–21.
- Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (3): 577–92.
- . 1997. "Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs." *Journal of Conflict Resolution* 41 (1): 68–90.
- Feng, Huiyun, Kai He, and Xiaojun Li. 2019. *How China Sees the World: Insights from China's International Relations Scholars*. Singapore: Palgrave Macmillan.
- Festinger, Leon. 2001. *A Theory of Cognitive Dissonance*. Reissued by Stanford Univ. Press in 1962, Renewed 1985 by author, [Nachdr.]. Stanford, Calif: Stanford Univ. Press.
- Fuhrmann, Matthew, and Todd S. Sechser. 2014. "Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence." *American Journal of Political Science* 58 (4): 919–35.
- Glaser, Charles. 1997. "The Security Dilemma Revisited" 50 (1): 171–201.
- Glaser, Charles L. 2010. *Rational Theory of International Politics the Logic of Competition and Cooperation*. Princeton, N.J.; Oxford: Princeton University Press.
- Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46 (2): 175–200.
- Herrmann, Richard K. 2017. "How Attachments to the Nation Shape Beliefs About the World: A Theory of Motivated Reasoning." *International Organization* 71 (S1): S61–84.
- Hu, Peng. 2018. "Popular Understanding of Democracy in Contemporary China." *Democratization* 25 (8): 1441–59.

- Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton, N.J: Princeton University Press.
- . 2006. “Understanding Beliefs.” *Political Psychology* 27 (5): 641–63.
- Katagiri, Azusa, and Eric Min. 2019. “The Credibility of Public and Private Signals: A Document-Based Approach.” *American Political Science Review* 113 (1): 156–72.
- Kertzer, Joshua D., and Ryan Brutger. 2016. “Decomposing Audience Costs: Bringing the Audience Back into Audience Cost Theory.” *American Journal of Political Science* 60 (1): 234–49.
- Kertzer, Joshua D., Ryan Brutger, and Kai Quek. 2024. “Perspective-Taking and Security Dilemma Thinking: Experimental Evidence from China and the United States.” *World Politics* 76 (2): 334–78.
- Kertzer, Joshua D., Brian C. Rathbun, and Nina Srinivasan Rathbun. 2020. “The Price of Peace: Motivated Reasoning and Costly Signaling in International Relations.” *International Organization* 74 (1): 95–118.
- Kertzer, Joshua D., and Jonathan Renshon. 2022. “Experiments and Surveys on Political Elites.” *Annual Review of Political Science* 25 (1): 529–50.
- Kertzer, Joshua D., Jonathan Renshon, and Keren Yarhi-Milo. 2021. “How Do Observers Assess Resolve?” *British Journal of Political Science* 51:308–30.
- Kim, Seok Joon. 2022. “Doom and Gloom, From Structure to Human Minds: What Makes a North Korean Nuclear Deal Difficult?” *Political Psychology* 43 (4): 715–30.
- Kurizaki, Shuhei, and Taehee Whang. 2015. “Detecting Audience Costs in International Disputes.” *International Organization* 69 (4): 949–80.

- Kydd, Andrew H. 1997. "Sheep in Sheep's Clothing: Why Security Seekers Do Not Fight Each Other." *Security Studies* 7 (1): 114–55.
- . 2007. *Trust and Mistrust in International Relations*. Third printing and First paperback printing. Princeton, NJ Oxford: Princeton University Press.
- Lacina, Bethany, and Charlotte Lee. 2013. "Culture Clash or Democratic Peace?: Results of a Survey Experiment on the Effect of Religious Culture and Regime Type on Foreign Policy Opinion Formation." *Foreign Policy Analysis* 9 (2): 143–70.
- Levendusky, Matthew, and Michael Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74 (2): 323–38.
- Levin, Dov H, and Robert F Trager. 2019. "Things You Can See From There You Can't See From Here: Blind Spots in the American Perspective in IR and Their Effects." *Journal of Global Security Studies* 4 (3): 345–57.
- Li, Xiaojun, and Dingding Chen. 2021. "Public Opinion, International Reputation, and Audience Costs in an Authoritarian Regime." *Conflict Management and Peace Science* 38 (5): 543–60.
- Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946–1986." *American Political Science Review* 87 (3): 624–38.
- McManus, Roseanne W. 2014. "Fighting Words: The Effectiveness of Statements of Resolve in International Conflict." *Journal of Peace Research* 51 (6): 726–40.
- McMorrow, R.W. 2015. "Membership in the Communist Party of China: Who Is Being Admitted and How?" *JSTOR Daily*, December 19, 2015.
<https://daily.jstor.org/communist-party-of-china/>.

- Nomikos, William G., and Nicholas Sambanis. 2019. "What Is the Mechanism Underlying Audience Costs? Incompetence, Belligerence, and Inconsistency." *Journal of Peace Research* 56 (4): 575–88.
- Peyton, Kyle, Gregory A. Huber, and Alexander Coppock. 2022. "The Generalizability of Online Experiments Conducted During the COVID-19 Pandemic." *Journal of Experimental Political Science* 9 (3): 379–94.
- Ramsay, Kristopher W. 2004. "Politics at the Water's Edge: Crisis Bargaining and Electoral Competition." *Journal of Conflict Resolution* 48 (4): 459–86.
- Reilly, James. 2012. *Strong Society, Smart State: The Rise of Public Opinion in China's Japan Policy*. Contemporary Asia in the World. New York: Columbia University Press.
- Renshon, Jonathan, Keren Yarhi-Milo, and Joshua D. Kertzer. 2023. "Democratic Reputations in Crises and War." *The Journal of Politics* 85 (1): 1–18.
- Risse-Kappen, Thomas. 1995. "Democratic Peace — Warlike Democracies?: A Social Constructivist Interpretation of the Liberal Argument." *European Journal of International Relations* 1 (4): 491–517.
- Sartori, Anne E. 2002. "The Might of the Pen: A Reputational Theory of Communication in International Disputes." *International Organization* 56 (1): 121–49.
- . 2005. *Deterrence by Diplomacy*. Princeton, N.J: Princeton University Press.
- Schafer, Mark, and Stephen G. Walker. 2006. "Democratic Leaders and the Democratic Peace: The Operational Codes of Tony Blair and Bill Clinton." *International Studies Quarterly* 50 (3): 561–83.
- Schultz, Kenneth A. 1998. "Domestic Opposition and Signaling in International Crises." *American Political Science Review* 92 (4): 829–44.

- Slantchev, Branislav L. 2005. "Military Coercion in Interstate Crises." *The American Political Science Review* 99 (4): 533–47.
- . 2006. "Politicians, the Media, and Domestic Audience Costs." *International Studies Quarterly* 50 (2): 445–77.
- Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (3): 623–38.
- Snyder, Jack, and Erica D. Borghard. 2011. "The Cost of Empty Threats: A Penny, Not a Pound." *American Political Science Review* 105 (3): 437–56.
- Ternovski, John, Lilla Orr, Joshua Kalla, and Peter Aronow. 2022. "A Note on Increases in Inattentive Online Survey-Takers Since 2020." *Journal of Quantitative Description: Digital Media* 2 (February).
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61 (04).
- Tomz, Michael R., and Jessica L. P. Weeks. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107 (4): 849–65.
- Tomz, Michael, Jessica L.P. Weeks, and Keren Yarhi-Milo. 2020. "Public Opinion and Decisions About Military Force in Democracies." *International Organization* 74 (1): 119–43.
- Trachtenberg, Marc. 2012. "Audience Costs: An Historical Analysis." *Security Studies* 21 (1): 3–42.
- Trager, Robert F. 2010. "Diplomatic Calculus in Anarchy: How Communication Matters." *The American Political Science Review* 104 (2): 347–68.

- Turcsanyi, Klara Dubravcikova, Kristina Kironska, Tao Wang, James Iocovozzi, Gries Gries, Veronika Vasekova, and Andrew Chubb. 2022. “Chinese Views of the World at the Time of the Russia-Ukraine War Evidence from a March 2022 Public Opinion Survey.” Palacky University Olomouc and CEIAS. <https://ceias.eu/chinese-views-of-the-world-at-the-time-of-the-russia-ukraine-war/>.
- Uzonyi, Gary, Mark Souva, and Sona N. Golder. 2012. “Domestic Institutions and Credible Signals: *Domestic Institutions and Credible Signals*.” *International Studies Quarterly* 56 (4): 765–76.
- Weeks, Jessica L. 2008. “Autocratic Audience Costs: Regime Type and Signaling Resolve.” *International Organization* 62 (01).
- Weiss, Jessica Chen. 2013. “Authoritarian Signaling, Mass Audiences, and Nationalist Protest in China.” *International Organization* 67 (1): 1–35.
- Wendt, Alexander. 1994. “Collective Identity Formation and the International State.” *American Political Science Review* 88 (2): 384–96.
- Yarhi-Milo, Keren. 2013. “In the Eye of the Beholder: How Leaders and Intelligence Communities Assess the Intentions of Adversaries.” *International Security* 38 (1): 7–51.
- Yarhi-Milo, Keren, Joshua D. Kertzer, and Jonathan Renshon. 2018. “Tying Hands, Sinking Costs, and Leader Attributes.” *Journal of Conflict Resolution* 62 (10): 2150–79.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*.