Observable Bounds of Rationality and Credibility in International Relations

Andrew Kenealy
PhD Candidate
Duke University
Department of Political Science
andrew.kenealy@duke.edu

Trent Ollerenshaw
PhD Candidate
Duke University
Department of Political Science
trent.ollerenshaw@duke.edu

So Jin Lee
Grand Strategy, Security, and Statecraft Postdoctoral Fellow
Harvard Kennedy School International Security Program
Massachusetts Institute of Technology Security Studies Program
sojinlee@hks.harvard.edu

Abstract: In crisis bargaining scenarios, resolved states send costly signals to demonstrate a willingness to fight. Yet public signals of resolve are issued by specific leaders, who operate with cognitive limitations in challenging decision environments. Drawing upon theories of behavioral economics, we develop a series of observable indicators plausibly connected to leader ability and difficulty of decision environment that we expect interactively shape perceived signaler credibility: capable leaders should be better able to overcome difficult decision environments. Effects, we argue, are due to perceived variation in the signaler's uncertainty over the future costs at stake. Preliminary analysis of a pilot conjoint experiment provides suggestive support for expectations. Existing scholarship on credibility in IR emphasizes the value of power and of appearing tough; if confirmed, our results demonstrate the importance of appearing reasoned and methodical—qualities that most policymakers already strive to publicly convey.

In a now famous press briefing in 2012, President Barack Obama threatened the use of military force against Syria. "We have been very clear to the Assad regime," Obama explained, "that a red line for us is, we start seeing a whole bunch of weapons moving around or being utilized." That, he added, "would change my calculus" (Landler 2012).

When it became clear roughly a year later that the Assad regime used chemical weapons, President Obama elected not to use force, and instead secured a diplomatic solution. Yet the backtrack kicked off a political firestorm. Obama later told a journalist that he recognized that backing down in Syria after creating the perception that his reputation was tied to the issue had hurt him politically (Goldberg 2016). The episode seems like one of the best examples in recent memory of the concept of audience costs at work: the president made a threat, and appeared to pay a price when he did not enforce it.

One possible assessment of Obama's deterrent statement suggests that it was a well-placed gamble, arising out of an accurate probabilistic judgement that the likelihood that it would deter the Bashar al-Assad regime, together with the value he associated with that outcome, outweighed the expected cost of backing down should deterrence fail (Chollet 2016). Perhaps Obama, then, was indeed acting optimally in light of his own interests, and would and should make the same strategic choices again if he had to. Perhaps he, in a probabilistic universe, simply got unlucky.

However, if President Obama—perhaps due to relative inexperience and the reportedly significant time pressures he faced as he formulated a response—mishandled the uncertainties involved, the narrative of his Syria red line reads quite differently. If Obama, for example, overestimated the likelihood that his deterrent statement would succeed, and underestimated how he would perceive the costs of military action if his line were crossed in the future due to personal biases and challenging circumstances, then issuing the statement was perhaps not his 'optimal'

strategy. Perhaps Obama did not predict the situation he would face in Syria and at home down the road, and in turn, did not foresee how he would judge those tradeoffs. If so, the signal he issued may have seemed like the optimal choice at the time. But with the benefit of hindsight, and in light of new information (that a president with more experience and reduced time constraints could plausibly have foreseen), Obama himself might view it as a misstep.

A vast and growing body of international relations (IR) scholarship examines when and why some publicly issued commitments, such as the threat made by Obama, are credible (Jervis, Yarhi-Milo, and Casler 2021).[1] Nevertheless, existing research tends to overlook the intuitive possibility that observers may make systematic judgments regarding whether signals themselves represent deviations from the leader's 'optimal' and 'rational' decision. In other words: whether signals appear to be mistakes.

The tendency to overlook the possibility of mistaken signals is perhaps all the more surprising considering we are regularly confronted with the (often tragic) reality that leaders are human, and therefore fallible. The signals that leaders send are the product of their efforts to make the best possible choices while operating under cognitive constraints and navigating challenging, fluctuating decision-environments. Indeed, Obama's Syria red line was reportedly unscripted, and followed a rushed series of crisis management meetings that were led by a remarkably intellectual, but comparatively inexperienced, first term president. Political scientists have long acknowledged that leaders, just like the rest of us, are 'boundedly rational', and that their decision-making is constrained by their cognitive capacities and environmental challenges (Bendor 2010; Holsti 1976;

[1] Following Jervis et al. (2021), we define credibility as "the extent to which an actor's statements or implicit commitments are believed" (169).

Rathbun 2019). We should expect, then, that elements of both leader skill and environmental difficulty shape the processes by which signals are produced. We might *also* expect, in turn, that plausibly observable indicators of cognitive and environmental constraints shape how signals are externally perceived.

Following from this basic observation, in this article we ask and strive to answer two related questions. Can third party observers identify 'boundedly rational' signaling behavior in leaders? And if they can, do these judgments shape their assessments of that signal's credibility? We present a theory that relates plausibly observable, boundedly rational signaling behavior to perceptions of credibility. The theoretical mechanism we propose is grounded in observer perceptions of signaler uncertainty. We argue that signals that appear to be the products of highly bounded (which we dub *constrained*) signaler behavior will be perceived to be the outcomes of decisions made under relatively greater uncertainty. These are thus noisy signals that do relatively little to reveal senders' future behavior to the observer. Why believe a leader's commitment when they seem like they might not anticipate, nor have internalized, the likely downstream costs at stake? Conversely, signals that seem produced by comparatively less bounded (or *unconstrained*) behavior should be judged by observers to be the outcomes of decisions made under relatively lesser uncertainty. When signalers appear to have mitigated more of their own uncertainty, their statements should have a greater capacity to inform observer beliefs about their future behavior.

Drawing upon insights from across the social sciences, we develop two sets of observable indicators of dispositional and environmental constraints that are theoretically connected to a leader's ability and the difficulty of the task environment, respectively. Dispositional constraints are traits and characteristics of the particular leader; environmental contraints relate to situational characteristics of the decision environment within which they operate. We argue that, together,

these indicators shape external perceptions of the overall decision-constraints upon leaders as they formulate a signal of commitment in IR. We expect signals that appear to be the product of unconstrained decision-processes will be judged to be more credible than those that appear highly constrained. Moreover, we expect that effects of leader ability and task environment on perceived credibility will be conditional on one another. More skilled leaders will be judged to be able to better overcome relatively challenging signaling environments than their less skilled counterparts. Just as we wouldn't anticipate chess masters to play tic-tac-toe much more effectively than the average adult, increased leader ability should be most impactful on assessments of credibility in international relations when the decision environment is particularly difficult (Bendor 2010).

To test these expectations, we adopt an experimental approach, which allows us to randomize sets of dispositional and environmental factors surrounding a hypothetical leader's signal of resolve, and draw clean inferences about observers' assessments of credibility. This registered report presents a design for a conjoint experiment that will allow us to parse the effects of dispositional and environmental indicators of decision-constraints, and additionally, to test their possible interaction. We present an analysis of pilot data, and specify planned analytic decisions.

Taken together, we hope that our theory and eventual findings will broaden scholars' understanding of credibility in international politics, and point to the existence of an alternative, complementary pathway to signaling resolve. Traditional models of credibility often emphasize the overall balance of power, as well as the importance of appearing tough and determined. Our account suggests the value of appearing reasoned, experienced, and methodical—characteristics that most policymakers already strive to emulate and publicly convey. It represents, to our knowledge, the first effort to systematically show how individual and environmental decision constraints jointly set the bounds of leader rationality, and in turn shape credibility in IR.

Costly Signaling Under Uncertainty

One of the central questions in the study of IR is how states discern the capabilities, interests, and resolve of allies and adversaries alike. Since at least some portion of resolve is private information (Kertzer 2016), and the international system incentivizes states to appear both stronger and more resolved than they are (Fearon 1995), states struggle to credibly convey their commitments.

IR theory tells us that states may overcome these 'type separation' problems by sending costly signals that others could not or would not send (Fearon 1997; Schelling 1966). One type of signal involves tying hands, including through issuing public statements, which entail "audience costs" that are paid ex post should the signaler break the stated commitment. Domestic audiences, the argument goes, value consistency and will punish leaders who abruptly change course. While some scholars raise a word of caution regarding the weight and prevalence of audience costs in the practice of international politics (Mercer 2012; Snyder and Borghard 2011; Trachtenberg 2012); others have experimentally demonstrated that these costs can, in fact, exist among domestic audiences (Tomz 2007; Tomz and Weeks 2013; Trager and Vavreck 2011) as well as international audiences (Kertzer, Renshon, and Yarhi-Milo 2019a), including when leaders find alternative solutions and take pains to justify them (Levendusky and Horowitz 2012; Lin-Greenberg 2019). Audiences are heterogeneous (Brutger and Kertzer 2018), tend to hold stable assessments of leaders across time (Croco, Hanmer, and McDonald 2021), and care about more than just consistency (Kertzer and Brutger 2016; Levy et al. 2015; Nomikos and Sambanis 2019), but most signs suggest the costs they impose do in fact lend credibility to leaders' public statements (Yarhi-Milo, Kertzer, and Renshon 2018). Although neither their centrality (Press 2005) nor their scope (Moon and Souva 2016) should be overstated, audience costs appear capable of underpinning at least some types of crisis bargaining interactions.

IR scholars have made remarkable progress identifying theoretically important, externally observable features and histories of the leader and state making a public commitment that might shape perceived credibility (Jervis, Yarhi-Milo, and Casler 2021). Yet signals that may generate audience costs are not issued in a vacuum—they are strategic choices made by idiosyncratic leaders attempting to secure their most preferred outcomes (Lupton 2020). Since we tend to agree that leaders of countries who must make crisis scenario decisions do not, in fact, operate with perfect information or perfect abilities to parse complex foreign policy decisions, conducting this required calculation in practice is difficult. Horowitz et al. (2018) suggest, for example, that leaders may vary in their beliefs regarding the efficacy of military power and risk tolerance based on their individual military backgrounds, and thus that some leaders may select into better types of threats over time.

Beleaguered leaders considering sending a signal of commitment during a crisis must complete a difficult task. They must project themselves into the future, and imagine how they will wrestle with a host of possible decisions, and how they will perceive possible future costs, including the costs of fighting and costs of backing down. All of these these future costs are characterized by varying degrees of uncertainty. Although uncertainty has meant different things to different groups of scholars of IR, we adopt here the rationalist approach (Rathbun 2007). Uncertainty, in this case, captures the range of an actor's future perceived costs. Importantly, uncertainty over these values is epistemic. That is, the future values are at least partially knowable, but in practice, impossible to perfectly resolve (Friedman 2019).

Reducing Uncertainty

Scholars appear to be reaching a near consensus that elites (as well as ordinary citizens) vary in their capacities to reduce epistemic uncertainty, particularly in the arena of international

affairs (Friedman 2019; Friedman and Zeckhauser 2018; Tetlock 2017). Biases abound, over-confidence looms, and miscommunications between decision-makers and advisors are ubiquitous. Even so, some individuals and teams manage to consistently outperform others in geopolitical forecasting tasks (Horowitz et al. 2019; Tetlock and Gardner 2016).

In the IR context, then, leaders that are better able to resolve epistemic uncertainty should also be better equipped to conduct the calculations required of them to navigate iterated crisis bargaining scenarios. They should, importantly, be less likely to discover in later crisis stages that the tradeoffs with which they must wrestle—and the perceived costs of fighting and costs of backing down—have shifted in major ways for which they did not previously anticipate. Over the course of a crisis, leaders will gain information, including regarding the form of adversarial response, that naturally resolves uncertainty over the ultimate costs of fighting and costs of backing down, given that they send a signal. But at early stages, much is unknown. Consider, for example, the leader who discovers that the costs of fighting appear much greater than they had previously initially anticipated. They may send a signal of commitment only to backtrack when these costs later become evident. Obama's red line may be one such illustrative example. The converse may also occur: a leader who initially makes concessions to avoid a fight may later discover that the costs of backing down are in fact much greater than expected.[2] Perhaps the best known plausible historical example is the infamous policy of appeasement of Hitler's Germany in the 1930s, which culminated in the Munich Agreement of 1938 that formally oversaw German annexation of major

[2] We make no claims regarding which direction of unexpected shift is more likely, nor does our theory rely on a unidirectional assumption. The contention here is simply that some leaders are better able to predict uncertain costs and calibrate their early stage decision-making accordingly.

portions of Czechoslovak territory. Scholars continue to question whether appeasement policies were motivated by ultimately painful miscalculations rife with wishful thinking and other biases, or whether these were 'rational' decisions considering the options available (Ripsman and Levy 2008; Treisman 2004). But central to these debates are assessments of whether decision-makers handled uncertainty in biased ways, or whether they, in a probabilistic universe, were just unlucky.

Leaders' capacity to make 'optimal' upstream decisions thus depends on their ability to resolve uncertainty over their downstream expectations. Whether they can accurately resolve future uncertain costs shapes their likelihood of making signaling decisions in crisis bargaining scenarios that maximize their perceived payoffs, whatever they may be.

Constrained Leaders

In crisis bargaining contexts, when is the effective resolution of uncertainty most likely? While rational choice theory is often the basis for many models of human behavior in the social sciences, human rationality is generally more constrained than neoclassical models assume. Behavioral economics amends classical and neoclassical economic theory in an attempt to achieve more realistic accounts of economic processes and decision-making, so as to better explain empirical findings that seem to regularly break classic rational actor assumptions (Jones 1999; Simon 1997). Bounded rationality incorporates the environmental constraints and cognitive limitations which impair individuals' ability to make perfectly rational choices in a neoclassical sense. Simon, who coined the term, argues that much of strategic human behavior can be explained using bounded rationality as a theoretical basis; some of the most influential research on decision-making of the later 20th century takes bounded rationality as a point of departure (Kahneman and Tversky 1979).

Bounded rationality theory provides useful insights in this context because it presents at least two sets of plausibly observable criteria that may interactively shape individual-level

behavior: overall leader ability and environmental difficulties. To understand when individuals' choices might deviate from classically rational expectations, the theory adopts the metaphor of a pair of scissors. One blade is the "cognitive limitations" of actual humans, which may depend in part on relevant experience and can vary among individuals; the second is the "structure of the environment," including available informational resources and the time available to be devoted to a task, which can also vary across scenarios (Gigerenzer and Selten 2001). Understanding human behavior, the argument goes, requires both blades—either alone is insufficient.

Bounded rationality theory also suggests that the effects of cognitive constraints on behavior depend on environmental constraints. One of Simon's primary contentions was that cognitive constraints only "show through" when the problem is sufficiently challenging (Bendor 2010, 17). When the problem is comparatively easy, decision-makers can adapt to the environment at hand, and are able to effectively optimize. To be sure, crisis bargaining interactions are always challenging. Nevertheless, some situations are more difficult than others, and in these situations, cognitive constraints should be increasingly relevant. Political scientists have long emphasized the importance of interactions between leaders' cognitive processes and decision environments in shaping foreign affairs choices, including during international crises. And an interactive cognitive-environmental relationship has been central to models of decision in IR for decades; Holsti (1976, 27) notes that many IR scholars since at least the middle of the 20th century have simply "assumed" it.

Determining what, exactly, is rationally 'optimal' behavior for a decision-maker depends on the decision-maker's long-term perceived costs, which in crisis bargaining scenarios is at least partially private information (Kertzer 2016). However, these perceived downstream costs are at least partially unknown, even to the signaler, as they are shrouded in epistemic uncertainty

(Friedman 2019). Bounded rationality predicts that some leaders may be better able to resolve this uncertainty over future perceived costs—and thereby reduce the distance between their 'intendedly rational' signaling behavior and purely rational, unconstrained behavior—as uncertainty is reduced.

Observing Constrained Decisions

If leaders' decision-making is indeed constrained, third party observers might be able to recognize these decision-constraints, and incorporate evaluations of these constraints into their judgments of others' behavior. In other words, observers do not need to assume that actors are playing an optimal strategy even while they recognize all actors are intendedly rational. Since observers seem to probabilistically infer how crisis bargaining actors perceive relevant costs based on observable factors (Kertzer, Renshon, and Yarhi-Milo 2019b), parse distinctions among elites (Lupton and Webb 2022), and pay especially close attention to leader characteristics during the crisis escalation phase (Goldfien and Joseph Forthcoming), we anticipate that observers will also make judgments regarding whether actors are making signaling choices that appear likely to be aligned with their most likely optimal behavior.[3] Assessing the possibility of mistakes among leaders would seem to be a natural task, and a broad literature in the social sciences suggests that observers are indeed able to evaluate communicative sources, including leaders, along dispositional and environmental

[3] Of course, observers calculating the credibility of a particular signal may deviate from the 'rational' assessment process we outline here. Different sorts of observers interpret signals in systematically different ways (Yarhi-Milo, Kertzer, and Renshon 2018). We leave questions of whether and how possible observer variation shapes evaluations of credibility for future research.

dimensions as they assess their credibility (Burgoon, Birk, and Pfau 1990; Krcmaric, Nelson, and Roberts 2020; Lupia and McCubbins 1998; Lupton and Webb 2022).

But to what specific individual and environmental indicators in interstate crises might observers turn? Drawing upon a range of literatures, we derive five dispositional factors that plausibly indicate the signaler's level of ability, as well as three environmental factors which may indicate the difficulty of the signaling environment faced. We expect that these, together, will interactively shape observer perceptions regarding a signaler's decision-constraints.

Dispositional Constraints

At least five dispositional leader traits may shape observer's perceptions of leader ability amidst crisis bargaining scenarios: time in office, prior foreign policy experience, prior military service, age, and prior academic performance. This list is not exhaustive, but rather presents a set of likely indicators for a latent variable capturing leaders' dispositional constraints. These traits were selected based on their theoretical importance in prior literatures and plausible connection to latent cognitive ability. We briefly justify each trait selection in turn.

Experience. Observers may expect that leaders' prior experiences shape their relevant knowledge regarding foreign affairs and conflict, and in turn their ability to resolve uncertainty in these contexts. Following the traits highlighted by prior scholarship, we focus specifically on three types of potentially important leader experiences: 1) time in office (Bak and Palmer 2010; Gaubatz 1991; Gelpi and Grieco 2001; Potter 2007; Smith and Spaniel 2019); 2) broader non-military foreign policy experiences (Saunders 2017); and 3) military service (Horowitz et al. 2018).

Age. While existing political science works tends to focus on the relationship between leader age and conflict initiation (Horowitz, McDermott, and Stam 2005; Horowitz, Stam, and Ellis 2015), we approach and present leader age as a proxy for relative cognitive ability. It is

commonly believed that "declines in performance on cognitive tasks that require one to quickly process or transform information to make a decision" follow normal aging (Murman 2015), such that the elderly in particular may struggle to incorporate significant volumes of information, project themselves into the future, and/or resolve epistemic uncertainty relative to more youthful peers. Specifically, we believe this cognitive decline may enter observer's calculus non-linearly as a leader reaches their eighties, when pronounced cognitive decline is increasingly commonplace.

Academic Performance. Finally, we contend that observers may view educational background and prior academic performance as a proxy for leader competence, mental acuity, and/or cognitive ability (Besley, Montalvo, and Reynal- Querol 2011). Rather than manipulating the level of degree achieved (Carnes and Lupu 2015; Dreher et al. 2009; Horowitz, Stam, and Ellis 2015) or varying the field of study (Flores, Lloyd, and Nooruddin 2013) or location of study (Nelson 2017), we present evidence regarding the leader's academic performance based on publicly available transcripts from their time as a student. Political candidates in democracies, for example, often release their academic transcripts, and they generally receive public scrutiny.

Environmental Contraints

Bounded rationality suggests that situational factors within a decision environment may shape leaders' capacity to resolve uncertainty and make optimal decisions (Mintz and DeRouen Jr 2010; Payne, Bettman, and Johnson 1993). To assess this possibility, we vary three observable situational factors which should be related to difficulty of the decision-making environment: the reliability of information accessible to the leader, and whether the leader is experiencing temporary fatigue or other stress which might impair their attention to the crisis at hand. Again, we do not view these environmental constraints as exhaustive, but instead as valid indicators of the cognitive difficulty or permissiveness of a given decision environment amidst crisis bargaining.

Informational Resources. Information, particularly about adversary's intentions and capabilities, is key to decision-making in the bounded rationality framework (Jones 1999). Some leaders may have access to more reliable information than others based on the externally perceived quality of the information provided by advisors or the intelligence agencies at their disposal. In the context of our experiment, we manipulate the stated reliability of the leaders' information due to varied intelligence agency reputation.

Stress and Fatigue. It is generally understood that temporary high levels of stress, including emotional anxiety related to familial well-being, lead to negative individual performance (Coelho, Hamburg, and Adams 1974; Dolan 2016; Janis and Mann 1977; Markwica 2018; Mintz and DeRouen Jr 2010), including for political leaders during foreign policy crises (George 1986). Fatigue, such as that experienced due to time zone shifts or illness, may also adjust the perceived difficulty of the environment (Hermann 1979). When faced with stressors or fatigue, our expectation is that leaders will be less sharply attentive to the crisis at hand, in turn affecting their ability to reach optimal decisions. We manipulate stress and fatigue, respectively, by prompting respondents to consider a recent injury to a close family member of the leader and by making respondents aware that the leader is either experiencing jet lag from recent travel, or recovering from a mild case of the flu or COVID-19. Since the outbreak of COVID-19, for example, infected leaders have regularly made headlines, and the cognitive impairments of COVID-19 infections are widely discussed. We acknowledge that fatigue and stress are not the most common or pronounced situational impediments to leaders' ability to devote attention to international crises. Indeed, leaders' time and attention are far more frequently diverted due to concerns related to domestic governance. However, we do not manipulate factors related to domestic governance because these factors not only shape environmental difficulty, but may directly institutionally constrain leaders'

capacity to use force. Stress and fatigue may be relatively weak treatments for situational impediments to attentiveness, but they benefit from being plausibly exogeneous to the option set leaders have during crisis bargaining.

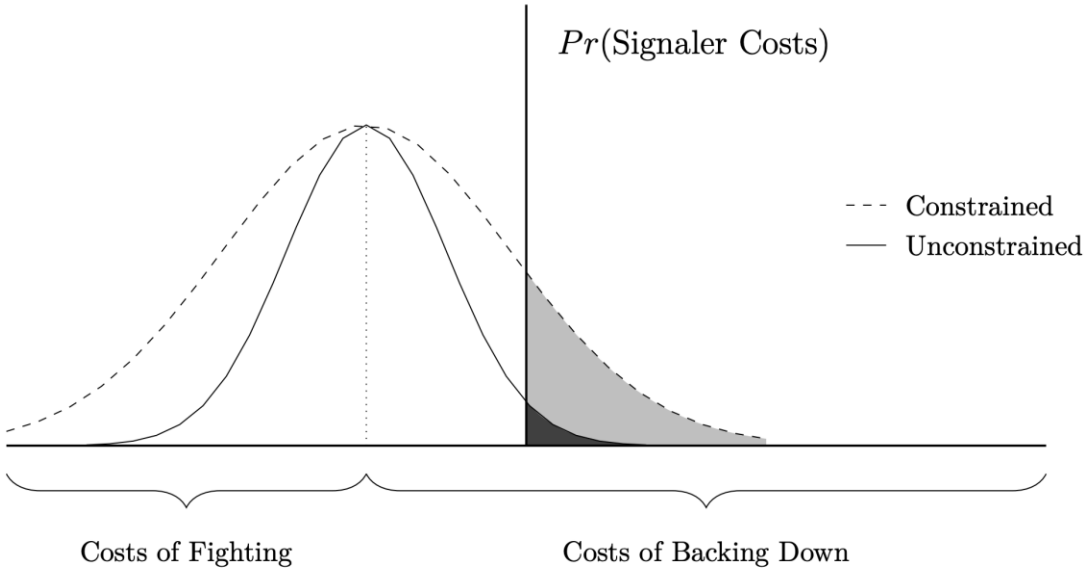
Relating Decision Constraints to Credibility

We argue that perceptions of decision constraints during signal formulation shape signal credibility because such perceptions shape observers' assessments of the quantity of uncertainty surrounding signalers' perceived future costs of fighting and costs of backing down. We contend that unconstrained signals can do more to reduce observer uncertainty over future signaler costs than comparatively constrained signals, and thus will be judged to be more credible. Highly constrained signals, on the other hand, do relatively little to reduce observers' uncertainty over how the signaler understands the relevant costs. We illustrate this argument with idealized depictions of how observers may make probabilistic judgments about how the signaler perceives costs.

The relevant calculus for the signaler when tested is whether their perceived costs of backing down exceed their perceived costs of fighting. If the perceived costs of backing down exceed the perceived costs of fighting, they will fight. If, on the other hand, the costs of fighting are ultimately perceived to be greater than the costs of backing down, the signaler will back down. External assessments of credibility—whether the signaler will fight if tested—thus hinge on assessments of the signaler's understanding of these costs. Because bluffing could in fact be an optimal strategy for some signalers, and these costs can never be perfectly revealed, no signals can achieve perfect credibility. Careful observers assessing credibility may, therefore, assign probabilities over the balance between the two values, since it is the balance between the costs of fighting and costs of backing down that ultimately determines what the signaler will do. We depict an idealized observer estimation of two different signaler's internally perceived costs in Figure 1.

First, consider the idealized observer assessment of the credibility of an unconstrained signaler, depicted in Figure 1 by a probability distribution drawn with a solid line. The observer estimates that this unconstrained signaler's costs of backing down are likely to be greater than the costs of fighting. The estimated most likely balance between the two relevant costs is represented by the peak of the distribution, which is at the midpoint, and marcated by the vertical dotted line. The white shaded region underneath the distribution captures the overall estimated probability that the costs of fighting will be less than the costs of backing down, and that the signaler will stand firm (and thus that the signal is credible); the black shaded region captures the overall estimated probability that costs of fighting will be greater than the costs of backing down, and that the signaler will back down. In this instance, as the black region is prorportionately quite small, the observer judges the signal very likely to be credible.

Figure 1: Observer Assessment of Signal Credibility



Yet, as we have argued, signals are products of a decision-process in which the prospective signaler wrestles with uncertain future costs. We expect that leaders in crisis bargaining scenarios who

have high levels of cognitive and/or environmental constraints will be judged less capable of resolving the uncertainty surrounding their future perceived costs of backing down and fighting. The signals that they send are thus more likely to be the outcomes of ex ante decision-processes characterized by greater levels of uncertainty over downstream costs. Constrained signalers' perceived costs of fighting and costs of backing down should, in turn, be more likely to shift farther over time ex post—and, crucially, in ways that they did not foresee—as the signaler gains information over the course of the crisis. The signals that such constrained decision-makers send, therefore, are noisy. If observers do in fact attend to the decision constraints that signalers face, they should judge signals sent by constrained signalers to be less reliable upstream indications of their senders' downstream perceived costs of fighting and backing down.

In response, we expect that observers will broaden their distributions over the signaler's perceived costs of fighting and costs of backing down to capture the correspondingly wider range of likely values. The idealized observer assessment of a constrained signaler is indicated by the dashed distribution in Figure 1. The consequence of this broader distribution is that more overall probability is assigned to the outcome in which the signaler backs down. This is reflected by the light shaded region of the distribution underneath the dashed line, which is proportionately larger than the dark shaded region under the solid line. This region captures the overall estimated probability that the signaler's perceived costs of backing down will ultimately be lower than their perceived costs of fighting, in which case they will back down. This, in turn, decreases observer assessments of signaler credibility, even though the most likely point estimate balance of costs (the peak of the distribution) is unchanged.

Figure 1 illustrates what amounts to, we think, an intuitive theoretical argument. We can conceive of a signal as an effort by a leader to reduce observers' uncertainty over their downstream

perceived costs. But for signals to meaningfully reduce some portion of observer uncertainty over how the signaler perceives costs, and in turn what the signaler will do, observers must judge that the signalers *themselves* have appropriately reduced uncertainty over their own estimated future costs of fighting and costs of backing down. Observers have little reason to believe signals when the signaler appears to have a poor grasp on how they may experience the downstream costs at stake and, in turn, little idea at the time of the signal of what they will ultimately want in the future.

Empirical Expectations

The theoretical argument generates the following empirical expectations. First, reduced dispositional constraints that suggest a higher leader ability will increase estimations of a signal's credibility (H1). Second, reduced environmental constraints that suggest a relatively easier decision environment will increase estimations of a signal's credibility (H2).

Table 1: Summary of H1 and H2

Constraint Type	Condition	Direction	Credibility
Dispositional	Years in Office	↑	↑
	Foreign Policy Experience	↑	↑
	Military Experience	↑	↑
	Age (over 80)	↓	↑
	Academic Performance	↑	↑
Environmental	Information / Intelligence	↑	↑
	Fatigue / Stress	↓	↑

Finally, following from bounded rationality theory, we expect the effects of dispositional and environmental conditions on perceived credibility will be conditional on one another. Cognitively constrained leaders will incur greater penalties to credibility when faced with higher environmental constraints than highly skilled leaders. In other words, when decision-makers operate in a very difficult environment, cognitive constraints will be more likely to bind, and leader skill will be a

substantively more important predictor of credibility in difficult environments than in permissive environments. Because we stipulate that dispositional and environmental factors interact, this is an 'interactionist' theoretical expectation. Specifically, we expect that the positive effects of increasing leader ability on signal credibility increase as environmental difficulty increases (H3).

Design

To test our expectations, we plan to field a conjoint experiment on Amazon MTurk via CloudResearch.[4] A conjoint experiment is the appropriate tool to test our argument for at least three reasons. First, an experimental approach allows us to isolate the causal effects of treatments on respondent perceptions of credibility, albeit in an abstract setting (Brutger et al. 2022). Second, a conjoint experiment allows us to simultaneously assess the impact of a range of dispositional and situational factors on perceptions of credibility, as well as consider how they may interact, while achieving necessary statistical power and avoiding respondent satisficing (Bansak et al. 2018). Finally, since research suggests that elite-public behavioral gaps rooted in specialized expertise are not particularly large (Kertzer 2020), we judge that our claims based on an analysis of a convenience sample are at least plausibly generalizable to elites. Indeed, to the extent that MTurk workers are capable of making the evaluations we are interested in, attentive elite actors should make even more well-considered evaluations when there are real stakes to credibility assessments.

Sample Size and Data Quality

[4] Other than risks from normal computer use, this study does not present any risks to human subjects, does not involve deception, and was reviewed and approved by the [insert university] Institutional Review Board ([insert Protocol #)].

We plan to survey 1,200 MTurk workers via CloudResearch, each of which will provide ten observations for our main dependent variable of interest via the conjoint design (i.e., N=12,000). Prior research finds respondents remain highly attentive to varied attributes of the number we propose (9) (Bansak et al. 2021) and for the number of repeated tasks we propose (10) (Bansak et al. 2018). A sample of this size would also provide at least 95% statistical power given our minimum effect sizes of interest (see Appendix C for a full discussion of minimum detectable effect sizes, sample size, and statistical power).

We will also take several steps to ensure data quality. We will collect a sample of MTurk workers who have completed 1000+ HITs with a HIT approval rate of 98% or more. We will use CloudResearch panel workers who are drawn from the pool of MTurk workers, but required to pass a series of attention and engagement checks to participate in CloudResearch panels. We will also block duplicate IP addresses, suspicious geocodes, and remove all respondents using VPNs that suggest they are international workers (workers with these characteristics tend to give lower quality responses; see Kennedy et al. 2020). At the start of the survey, respondents will be required to complete a captcha-style question. Respondents who fail to correctly identify the two images with stop signs, and only those two images, will be immediately removed from the survey. About a minute into the survey, respondents will be asked a factual question about the non-experimental scenario we ask them to read (Kane and Barabas 2019). Respondents who fail to correctly answer this factual question will be removed from the survey. All criteria for which respondents could be excluded from our sample are based on pre-treatment variables to ensure we are collecting an attentive sample without conditioning sample inclusion on post-treatment variables (Kane, Velez, and Barabas 2022; Montgomery, Nyhan, and Torres 2018). Non-qualifying respondents will *not* count towards our intended 1,200 respondent sample, nor will non-complete survey responses (i.e.,

respondents who do not reach the last screen of the survey), but partially complete respondents who skip one or more questions (which we found to be rare in our pilot study) will be counted towards the planned 1,200 respondent sample.

Experimental Procedures

The intended experimental design is as follows. After consenting to participate in the survey, respondents are provided an instructional briefing of what to expect. This briefing includes visual examples and explanations of each of the aforementioned dispositional and situational manipulations. Respondents are shown a vignette describing a scenario between two hypothetical countries, Country A and Country B, engaged in a dispute over territory along their shared border. They are told next that, in light of Country B's aggression, the leader of Country A: "must consider their options and make a decision about how they will respond to Country B. They must choose whether to appear committed OR to concede and avoid a fight." Respondents are told that the details of this scenario will never change over the course of the survey.

Respondents are then asked to imagine a hypothetical leader of Country A, and are shown a profile of a sample leader experiencing different decision environments. A sample conjoint profile is shown in Table 2. Randomization of the order of "Leader Profile" and "Other Situational Details" blocks and attributes' order within the two blocks will occur at the respondent level.

Table 2: Sample Conjoint Profile

Leader Profile	
Age	55 years old
Time as Leader	7 years
Foreign Affairs Experience	Yes
Military Experience	No
Academic Records	Very good
More Situation Details	
Intel on Country B	Very reliable
Fatigue	Jet lagged from international travel

Stress	None
Military Strength	Slightly stronger

Because respondents may draw inferences about a nation's military strength—thought to be a core feature of credibility—based on leaders' attributes or intelligence capabilities in ways that could bias estimations in our favor, we also include a relative military strength manipulation in the profile. This fixes respondents' estimations of national-level military strength for each profile, and also provides a benchmark against which we can weigh the estimated effects of our manipulations of interest.[5] But to clarify: military strength is *not* an environmental factor of interest with regards to the situational difficulties of parsing uncertainty.

Finally, after reading the leader profile, respondents are told of two new developments to the scenario. They are told, first, that the leader of Country A went on public TV and announced that they would "do whatever it takes" to defend the disputed territory from Country B. This constitutes a public signal of commitment which, if later broken, would incur audience costs. Second, they are told that Country B invaded two weeks later. The leader of Country A now faces a stark choice: they have to decide whether to "stand firm and defend the disputed territory" OR "back down and concede the territory to avoid a fight." After reading the full scenario, respondents are asked one question regarding leader credibility, which constitutes our main dependent variable. Respondents are asked to evaluate how likely they think each leader is to stand firm on a six-point scale recoded continuously to take values from 0 (extremely unlikely) to 1 (extremely likely).

[5] In the pilot, we included the military strength of Country A in the scenario, but believe that this information should be included in the table since respondents are likely more attentive to features in the changing conjoint tables than those of the fixed scenario over the course of the survey.

In subsequent iterations, respondents are presented with the same scenario, but a different table with newly generated leader profiles and situational factors to consider. Not including the 'practice' round, this exercise will be repeated a total of ten times per respondent. Each of the five dispositional and three situational factors (as well as national military strength) are randomized in each round. The probabilities for the levels of each manipulated attribute are shown in Table 3.

Table 3. Attribute Levels and Associated Probabilities of Selection

Attribute	*Condition 1*	*Condition 2*	*Condition 3*	*Condition 4*	*Condition 5*	*Condition 6*
Age	50-59 (5% each)	80-89 (5% each)	-	-	-	-
Leader Exp.	4-8 years (10 each%)	2-6 months (10% each)	-	-	-	-
Foreign Exp.	Yes (50%)	No (50%)	-	-	-	-
Military Exp.	Yes (50%)	No (50%)	-	-	-	-
Academic Records	Excellent (~16.7%)	Very Good (~16.7%)	Good (~16.7%)	Poor (~16.7%)	Very Poor (~16.7%)	Terrible (~16.7%)
Intel on Country B	Extremely Reliable (~16.7%)	Very Reliable (~16.7%)	Somewhat Reliable (~16.7%)	Somewhat Unreliable (~16.7%)	Very Unreliable (~16.7%)	Extremely Unreliable (~16.7%)
Fatigue	Jet lagged from international travel (25%)	Recovering from a mild case of flu (12.5%)	Recovering from a mild case of COVID-19 (12.5%)	No Fatigue (50%)	-	-
Stress	Leader's spouse recently hospitalized (12.5%)	Leader's child recently hospitalized (12.5%)	Leader's spouse recently injured in accident (12.5%)	Leader's child recently injured in accident (12.5%)	No Other Stress (50%)	-
Military Strength	Much Stronger (~16.7%)	Moderately Stronger (~16.7%)	Slightly Stronger (~16.7%)	Slightly Weaker (~16.7%)	Moderately Weaker (~16.7%)	Much Weaker (~16.7%)

After respondents complete ten rounds, we will present respondents with the same scenario and a newly generated profile, but we will ask two *different* questions. Respondents will be asked to evaluate the skill of the leader, as well as the difficulty of the situation the leader faced when

deciding whether to signal commitment. These questions will constitute manipulation checks, which we ask at the conclusion of the experiment to avoid priming respondents to consider leader skill or environmental difficulty as they evaluate credibility. This process will be repeated two additional times, for a total of three manipulation check observations per respondent (N=3,600).

Finally, we will collect information about respondents' political beliefs and demographic background for the purpose of reporting details about our sample's composition. Then, we will debrief respondents about their participation in the experiment.

Analyses

This section describes our intended statistical analyses and presents near-identical analyses from a pilot study. Note that we do not take the analyses that follow to provide evidence in support of or against our hypotheses—all are exploratory, and we made a number of analytic choices that warrant a hands-tied replication. We present the pilot results only to illustrate the analyses we intend to conduct with a new sample collected with the proposed design, and how we would present the results in relation to our theoretical expectations. These analytic choices and expectations are summarized in a Design Table presented in Appendix B. The intended data pre-processing and analytic pipeline in R is included in the supplementary materials, though we outline core features of the analysis here using the pilot study as an example where possible.[6] The pilot study includes a sample of 250 respondents recruited from Amazon MTurk via CloudResearch.[7] All but one respondent evaluated all ten profiles (they completed four), leaving us with a pilot

[6] The lone exception is the marginal effects plot (Figure 4) which is constructed in Stata using an identical multiplicative interaction model to that which can be produced using the R script.

[7] We paid respondents $1.70 for finishing the study, and plan to pay this rate for the full study.

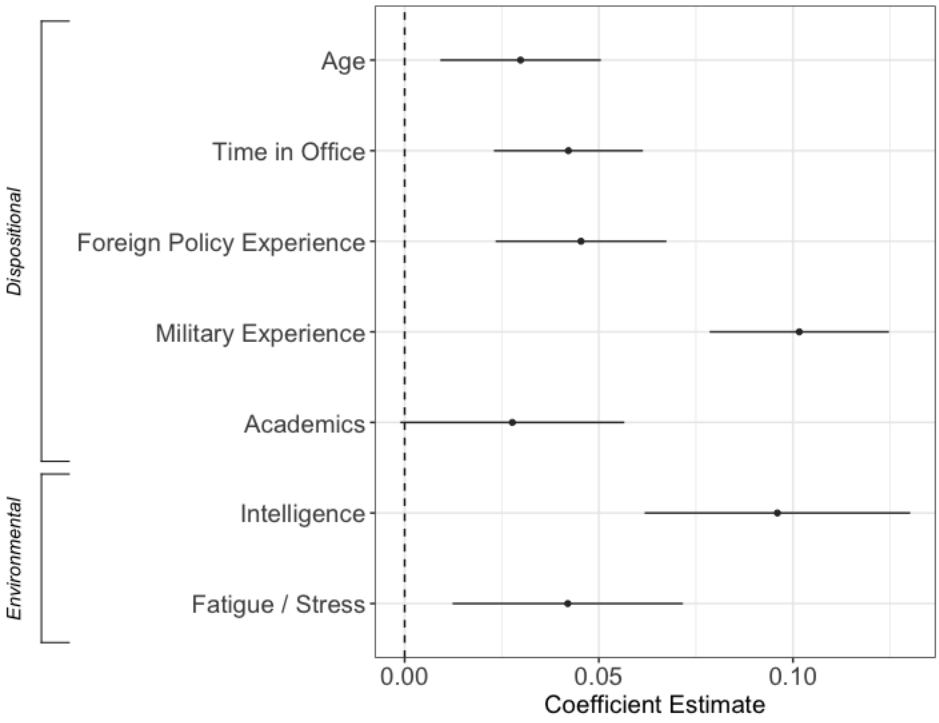
sample of 2,494 observations for analysis. Two things to note about the pilot are that it did not include the proposed factual manipulation check to screen out inattentive respondents, nor did it include relative military strength in the conjoint tables of attributes (instead, this information was embedded in the scenario itself). While the pilot study is therefore non-identical to the proposed full study, it follows most of our proposed experimental design.

Leader Ability and Situational Difficulty Component Effects

To test our hypotheses about the effects of dispositional (H1) and environmental (H2) attributes on perceived signal credibility, we estimate the Average Marginal Component Effects (AMCE) of each attribute. Most of the attributes are dichotomized when calculating AMCEs; for example, age can take values between 50-59 and 80-89, but is dichotomized to take values of 0 (80-89) or 1 (50-59) to reflect the fact that we expect large differences across, but not within, decades of age. Leader experience, foreign policy experience, and military experience are similarly dichotomized such that 0 indicates no experience and 1 indicates experience. Academic records, intelligence on Country B, and military strength, by contrast, are coded continuously from 0 to 1 where higher values reflect better leader abilities or easier situations since we assume the treatment levels of these attributes are approximately intervals. Finally, the different levels of stress and fatigue are treated as functionally equivalent (e.g., being fatigued from jetlag is treated as equivalent to being fatigued due to mild illness) and combined into a three-level variable that takes the values of 0 (stressed and fatigued), 0.5 (stressed or fatigued), and 1 (neither stressed nor fatigued). We judge this coding scheme for stress and fatigue appropropriately reflects the theoretical similarity of the stress and fatigue manipulations, which are split only so that we can include up to two sources of stress/fatigue in the tables without sacrificing readability.

In Figure 2, we display the AMCEs of each dispositional and situational attribute with 95% confidence intervals. AMCEs are estimated using a respondent fixed-effects model with robust standard errors clustered at the respondent level. Each AMCE represents the marginal effect of a given situational or dispositional manipulation averaged over the joint distribution of the other attributes (Hainmueller, Hopkins, and Yamamoto 2014). All attributes are coded such that positive AMCEs indicate support for our theoretical expectations. All p-values described for the pilot are derived using two-tailed tests since these analyses were exploratory; however, our confirmatory tests for the main study will use one-tailed (i.e., directional) hypothesis tests. Tables presenting results can be found in Appendix D.

Figure 2: Effects of Dispositional and Environmental Factors on Credibility



Looking first at the leader dispositional manipulations, we find that leaders do vary in perceived credibility as a function of background characteristics and relevant domain-specific experiences. We see statistically significant estimated effects for four of five attributes. Being non-

elderly and having domain-specific professional experiences (serving longer as leader, foreign policy experience, and military experience) increase the perceived credibility of the leaders' commitments to standing firm (all $p<0.01$). The one leader attribute that does not emerge as statistically significant is academic records, which exerts a non-significant 0.028-point effect in the expected direction ($p=0.060$). Were these confirmatory hypothesis tests, we would take the first four of them as offering strong evidence in support of our hypothesis that as leaders' domain-specific experiences and perceived mental acumen increase, so does their credibility during crisis bargaining scenarios.

Turning to the two environmental variables of interest, we also see statistically significant effects. The presence of stressors and fatigue, which we argue might temporarily impair a leader's ability to devote time or mental attention to the crisis at hand, are found to negatively impact leader credibility. The AMCE of having neither stress nor fatigue against a baseline of both stress and fatigue is significant and positive ($\beta=0.042$, $p=0.006$), indicating that stressors and fatigue may negatively impact leaders' credibility when signaling amidst crisis-bargaining scenarios. We also find that the reliability of a leader's information/intelligence when making their decision regarding whether to send a signal has an especially strong effect. Moving from having extremely unreliable to highly reliable information about the crisis at hand is associated with a 0.096-point increase in signal credibility ($p<0.001$). In general, these results suggest that costly signals sent by leaders operating under more difficult cognitive conditions are viewed as less credible than signals sent by leaders operating in permissive decision environments.

Overall, our pilot analysis aligns with our general expectations regarding the relationships between dispositional leader traits and situational difficulties of signaling games and perceptions of leader credibility. Leaders with traits that suggest stronger abilitites to think and reason through

a strategic design are perceived as more credible, and therefore more committed, than their peers who lack such traits; similarly, leaders who find themselves in cognitively permissive decision environments are perceived as being more credible when they send a signal than leaders who send signals in cognitively difficult situations.

The Interaction of Leader Ability and Enivironmental Difficulty

Having tested whether leaders' ability and the environments they operate in independently shape signal credibility, we next test H3: that low ability leaders' signal credibility is especially penalized relative to high ability leaders' signal credibility when making decisions in difficult environments. To test this hypothesis, we first construct additive indices to score leaders' dispositional abilities and environmental cognitive difficulties across all relevant attributes. This approach allows us to assess how the effect of aggregate leader ability may interact with the effect of aggregate environmental difficulty in a straightforward fashion. Testing interactions between each dispositional and environmental factor individually is neither theoretically appropriate nor empirically viable.

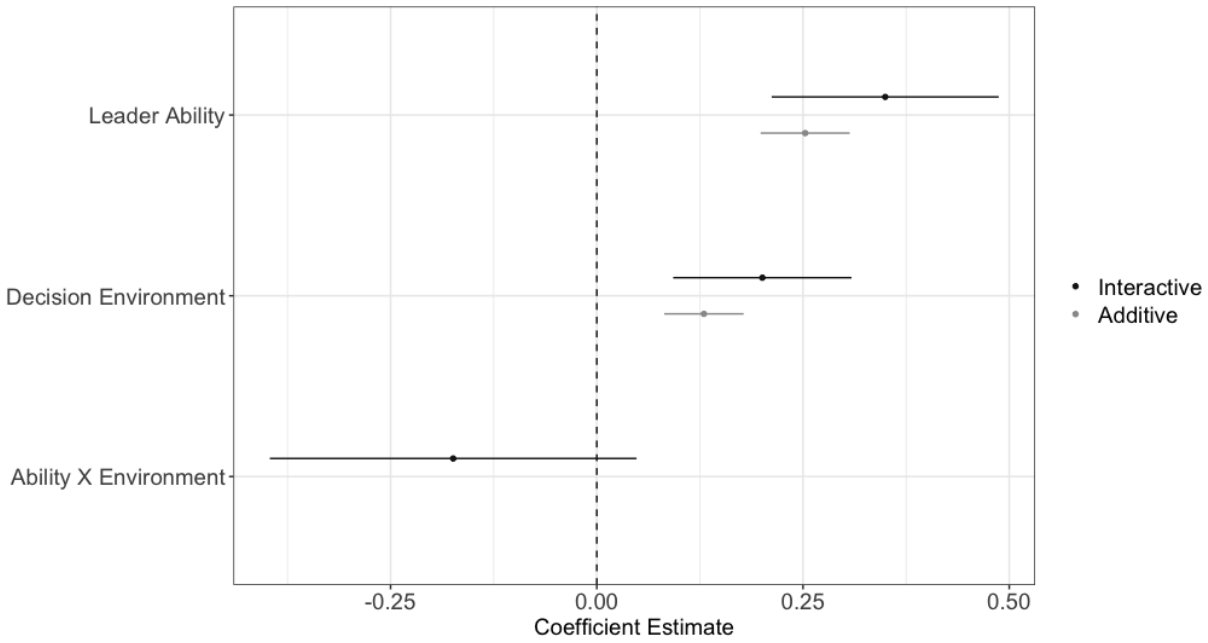
To contruct the leader ability index, we add the five leader attributes (age, time served as leader, foreign policy experience, military experience, academic records) and divide by five to create a measure that ranges from 0 (lowest ability) to 1 (highest ability). Similarly, we add the two environmental attributes (stress/fatigue and intel on Country B) and divide by two to create a measure of environmental difficulty that ranges from 0 (hardest environment) to 1 (easiest environment).[8] Given the coding of these indices, we expect a *negative* interaction effect; that is,

[8] Without strong priors about any particular attributes being more important than others, we opt to simply weigh all of the factors equally when constructing these indices.

we expect environmental difficulty to incur less of a penalty for credibility when the leader is viewed as high in ability or, conversely, that there will be less of a premium placed on being a high ability leader when the decision environment is very permissive.

We estimate two OLS models. The first regresses credibility on both indices additively; the second regresses credibility on both indices and includes the interaction between them. Again, our expectation is that both index coefficients will be positive in the non-multiplicative model, but that these indices will negatively interact in the multiplicative model because as environmental ease increases, the effects of increasing leader ability should decrease (and vice versa). As above, models are estimated using respondent-fixed effects with robust standard errors clustered at the respondent level. Coefficient estimates are plotted in Figure 3 with 95% confidence intervals, and p-values are derived using two-tailed tests since these analyses were exploratory, whereas our preregistered confirmatory hypothesis tests will use one-tailed (i.e., directional) tests.

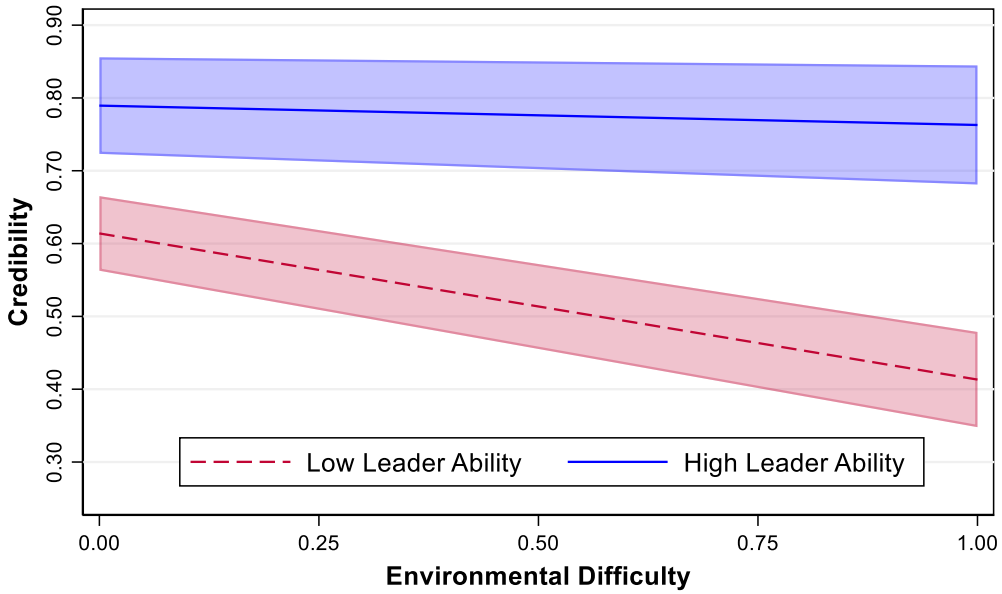
Figure 3: Effects of Leader Ability and Decision Environment on Credibility



We see from Figure 3 that, in the additive model, the estimated coefficients for the leader ability and decision environment indices are positive and statistically significant. This, of course, logically follows from the fact that the components of the indices were positively associated with credibility (see Figure 2). In the interactive model, we find a -0.174-point interaction effect that is in the expected negative direction, but is statistically indistinguishable from no interaction effect (p=0.125). For a table presenting model results, see Appendix D.

To aid in interpretation of the interaction between leader ability and decision environment, Figure 4 plots the predicted marginal effect of increasing environmental difficulty on signal credibility for the lowest and highest ability leaders (i.e., when leader ability is set to 0 and 1, respectively). For a more intuitive interpretation of the marginal effects, we reverse code decision environment difficulty such that a higher value indicates increasing environmental difficulty.

Figure 4: Effect of Environmental Difficulty by Leader Ability



As shown in Figure 4, high ability leaders are viewed as significantly more credible than low ability leaders at all levels of decision environment difficulty. High ability leaders are also viewed as being consistently credible on average, regardless of the environmental difficulties they

face. The predicted credibility of a high ability leader in the easiest environment (0.79) is near-identical to that of a high ability leader in the hardest environment (0.76)—an insignificant -0.03 effect of environmental difficulty on credibility ($p=0.697$). Low ability leaders, on the other hand, are viewed on average as credible in the easiest environments (0.61), but slightly non-credible in the most difficult environments (0.41). The -0.20 effect of environmental difficulty for low ability leaders is statistically significant ($p<0.001$). These results suggest that the penalty that comes with signaling in a cognitively challenging decision environment is borne primarily by leaders viewed as having a poor ability to parse uncertainty. However, in the pilot, the difference in differences between the effects of environmental difficulty by leader ability is non-significant ($\beta=0.174$, $p=0.125$), so we cannot definitively say that the effects of environmental difficulty truly differ by leader ability. We note, however, that our pilot was underpowered to detect an interaction effect of -0.174 points (or even effects larger than this in magnitude; see Appendix C), especially using two-tailed tests; an adequate test of the interaction hypothesis requires a much larger sample.

Conclusion

An enduring question in the study of international relations is why some threats can deter, while others cannot. Both classic and contemporary explanations largely focus on the cost of the signal to the sender, and attributes of the sender and state that indicate how that cost may be perceived. Drawing upon foundational scholarship in the social sciences, as well as recent advances in the behavioral study of IR, this article presents a novel theory of credibility predicated on observer assessments of a leader's decision constraints—the observable bounds of their rationality.

We argued that some leaders are better able to resolve the uncertainty inherent in crisis bargaining scenarios than others and, crucially, that observable features about the leader and the decision environment that they face indicate at least part of this variation. Our theory explains

why these observer perceptions of signaler uncertainty should shape credibility. Why believe a commitment from a leader who seems like they have little grasp on the future costs and stakes?

Preliminary analyses of pilot data suggest that observers calculating the credibility of a signal do indeed look to these dispositional and situational features as they make their judgments. Aligning with our expectations, multiple indicators of leader ability and of a permissive decision environment appear to increase perceptions of signal credibility. And while the estimated interaction effect is not statistically significant, the direction of the effect nonetheless provides some preliminary reason to believe that leader ability may be especially important as environmental difficulty increases. This hypothesis deserves being retested with a larger sample.

We believe a full experiment would be instructive no matter whether we can reject the null hypotheses. If our expectations are confirmed, our findings would present a number of implications for IR theory. First, they would suggest that idiosyncratic, temporary features of a signaler's decision environment, such as personal stresses, can shape observers' judgments regarding a given signal's credibility. These factors, in many cases, have no plausible connection to the overarching strategic balance between a signaler and their adversary, or how the signaler would internally perceive the reputational costs of backing down—which classical explanations suggest are principal drivers of credibility. Moreover, factors such as intelligence agency reputation—an understudied state-level, situational factor—may be shown to shape perceived credibility in more important ways than some dispositional factors, such as experience in office, that have received greater attention in the existing literature.

Confirmed effects for our manipulated dispositional factors, too, would broaden our understanding of the sorts of leader attributes that are relevant. We would confirm prior research that suggests that military experience affords special deterrent credibility, but would also show

that domain-specific expertise (no matter how that is acquired) matters as well. The relatively surprising impact of age further supports our expectation that perceptions of cognitive ability are central to perceptions of credibility.

In addition, evidence of interactions between dispositional characteristics and environmental difficulty would show that being perceived as more capable is especially valuable in relatively more challenging decision-environments. Such an interaction would provide important support for our proposed theoretical mechanism—that cognitive and environmental constraints jointly determine the decision constraints of a signaler, which in turn shapes credibility.

Perhaps most importantly, confirmatory results would uncover a new pathway to signalling credibility and deterrence. Prior research generally emphasizes the importance of being strong and seeming tough. Our findings, on the other hand, would suggest that it is also important to appear analytical, experienced, and measured. Careful crisis-management decision processes lend credibility to the signals they produce.

If our expectations are *not* confirmed, results would also be informative. In light of the steps we will take to ensure high data quality, we should have confidence that failure to reject the null hypotheses, or especially substantively weak results, are not false negatives. With the important caveat that even attentive survey respondents are not political elites, null results might reaffirm classic models of credibility in IR that emphasize the overall balance of power between nation states (Press 2005; Schelling 1966), the costs of reneging on a threat (Fearon 1997), as well as indicators of how leaders perceive those costs (Kertzer, Renshon, and Yarhi-Milo 2019a). Our inclusion of national-level relative military strength as a manipulation will help us parse the relative importance of classic explanations focusing on military power, with our own. Ultimately, if indicators of leaders' decision constraints—and the interaction between individual and

environment—prove unimportant to evaluations of credibility, this is new knowledge in the study of IR.

References

Bak, Daehee, and Glenn Palmer. 2010. "Testing the Biden Hypotheses: Leader Tenure, Age, and International Conflict." *Foreign Policy Analysis* 6(3): 257–73.

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. 2018. "The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments." *Political Analysis* 26(1): 112–19.

———. 2021. "Beyond the Breaking Point? Survey Satisficing in Conjoint Experiments." *Political Science Research and Methods* 9(1): 53–71.

Baranger, David A. et al. 2022. "Tutorial: Power Analyses for Interaction Effects in Cross-Sectional Regressions." https://psyarxiv.com/5ptd7/ (January 19, 2023).

Bendor, Jonathan. 2010. *Bounded Rationality and Politics*. Berkeley, CA: University of California Press.

Besley, Timothy, Jose G. Montalvo, and Marta Reynal- Querol. 2011. "Do Educated Leaders Matter?" *The Economic Journal* 121(554): F205-227.

Brutger, Ryan et al. 2022. "Abstraction and Detail in Experimental Design." *American Journal of Political Science*: ajps.12710.

Brutger, Ryan, and Joshua D. Kertzer. 2018. "A Dispositional Theory of Reputation Costs." *International Organization* 72(3): 693–724.

Burgoon, Judee K., Thomas Birk, and Michael Pfau. 1990. "Nonverbal Behaviors, Persuasion, and Credibility." *Human Communication Research* 17(1): 140–69.

Carnes, Nicholas, and Noam Lupu. 2015. "What Good Is a College Degree? Education and Leader Quality Reconsidered." *The Journal of Politics* 78(1): 35–49.

Chollet, Derek. 2016. *The Long Game: How Obama Defied Washington and Redefined America's Role in the World*. New York, NY: Public Affairs.

Coelho, George V., David A. Hamburg, and John E. Adams. 1974. *Coping and Adaptation*. Oxford, England: Basic Books.

Croco, Sarah E., Michael J. Hanmer, and Jared A. McDonald. 2021. "At What Cost? Reexamining Audience Costs in Realistic Settings." *The Journal of Politics* 83(1): 8–22.

Dolan, Thomas M. 2016. “Emotion and Strategic Learning in War.” *Foreign Policy Analysis* 12(4): 571–90.

Dreher, Axel, Michael J. Lamla, Sarah M. Lein, and Frank Somogyi. 2009. “The Impact of Political Leaders’ Profession and Education on Reforms.” *Journal of Comparative Economics* 37(1): 169–93.

Fearon, James D. 1995. “Rationalist Explanations for War.” *International Organization* 49(3): 379–414.

———. 1997. “Signaling Foreign Policy Interests: Tying Hands versus Sinking Costs.” *Journal of Conflict Resolution* 41(1): 68–90.

Flores, Thomas Edward, Gabriella Lloyd, and Irfan Nooruddin. 2013. “The Technocratic Advantage? Leadership and IMF Programs.” In *Prepared for Presentation at the 2013 Annual Convention of the International Studies Association, San Francisco, CA,*.

Friedman, Jeffrey A. 2019. *War and Chance: Assessing Uncertainty in International Politics*. New York, NY: Oxford University Press.

Friedman, Jeffrey A., and Richard Zeckhauser. 2018. “Analytic Confidence and Political Decision-Making: Theoretical Principles and Experimental Evidence From National Security Professionals.” *Political Psychology* 39(5): 1069–87.

Gaubatz, Kurt Taylor. 1991. “Election Cycles and War.” *Journal of Conflict Resolution* 35(2): 212–44.

Gelpi, Christopher, and Joseph M. Grieco. 2001. “Attracting Trouble: Democracy, Leadership Tenure, and the Targeting of Militarized Challenges, 1918-1992.” *Journal of Conflict Resolution* 45(6): 794–817.

George, Alexander. 1986. “The Impact of Crisis-Induced Stress on Decision Making.” In *The Medical Implications of Nuclear War*, Washington, DC: National Academy Press, 529–52.

Gigerenzer, Gerd, and Reinhard Selten. 2001. “Rethinking Rationality.” In *Rethinking Rationality: The Adaptive Toolbox*, Cambridge, MA: The MIT Press, 1–12.

Goldberg, Jeffrey. 2016. “The Obama Doctrine.” *The Atlantic*. https://www.theatlantic.com/magazine/archive/2016/04/the-obama-doctrine/471525/.

Goldfien, Michael, and Michael Joseph. Forthcoming. “Perceptions of Leader Importance in Conflict and Diplomacy: Evidence from the CIA’s President’s Daily Brief.” *Security Studies*.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments.” *Political Analysis* 22(1): 1–30.

Hermann, Margaret G. 1979. “Indicators of Stress in Policymakers during Foreign Policy Crises.” *Political Psychology* 1(1): 27–46.

Holsti, Ole R. 1976. “Cognitive Process Approaches to Decision-Making: Foreign Policy Actors Viewed Psychologically.” *American Behavioral Scientist* 20(1): 11–32.

Horowitz, Michael C. et al. 2019. “What Makes Foreign Policy Teams Tick: Explaining Variation in Group Performance at Geopolitical Forecasting.” *The Journal of Politics* 81(4): 1388–1404.

Horowitz, Michael C., Rose McDermott, and Allan C. Stam. 2005. “Leader Age, Regime Type, and Violent International Relations.” *Journal of Conflict Resolution* 49(5): 661–85.

Horowitz, Michael C., Philip Potter, Todd S. Sechser, and Allan C. Stam. 2018. “Sizing Up the Adversary: Leader Attributes and Coercion in International Conflict.” *Journal of Conflict Resolution* 62(10): 2180–2204.

Horowitz, Michael C., Allan C. Stam, and Cali M. Ellis. 2015. *Why Leaders Fight*. Cambridge University Press.

Janis, Irving L., and Leon Mann. 1977. *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment*. New York, NY, US: Free Press.

Jervis, Robert, Keren Yarhi-Milo, and Don Casler. 2021. “Redefining the Debate Over Reputation and Credibility in International Security: Promises and Limits of New Scholarship.” *World Politics* 73(1): 167–203.

Jones, Bryan D. 1999. “Bounded Rationality.” *Annual Review of Political Science* 2(1): 297–321.

Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47(2): 263–91.

Kane, John V., and Jason Barabas. 2019. “No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments.” *American Journal of Political Science* 63(1): 234–49.

Kane, John V., Yamil R. Velez, and Jason Barabas. 2022. “Analyze the Attentive & Bypass Bias: Mock Vignette Checks in Survey Experiments.”

Kennedy, Ryan et al. 2020. “The Shape of and Solutions to the MTurk Quality Crisis.” *Political Science Research and Methods* 8(4): 614–29.

Kertzer, Joshua D. 2016. *Resolve in International Politics*. Princeton, NJ: Princeton University Press.

———. 2020. “Re- Assessing Elite- Public Gaps in Political Behavior.” *American Journal of Political Science*: ajps.12583.

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, Jonathan Renshon, and Keren Yarhi-Milo. 2019a. "How Do Observers Assess Resolve?" *British Journal of Political Science*: 1–23.

Kertzer, Joshua D., Jonathan Renshon, and Keren Yarhi-Milo. 2019b. "How Do Observers Assess Resolve?" *British Journal of Political Science*: 1–23.

Kertzer, Joshua D, Jonathan Renshon, and Keren Yarhi-Milo. "Are Red Lines Red Herrings?" : 52.

Krcmaric, Daniel, Stephen C. Nelson, and Andrew Roberts. 2020. "Studying Leaders and Elites: The Personal Biography Approach." *Annual Review of Political Science* 23(1): 133–51.

Landler, Mark. 2012. "Obama Threatens Force Against Syria." *The New York Times*. https://www.nytimes.com/2012/08/21/world/middleeast/obama-threatens-force-against-syria.html (March 14, 2021).

Levendusky, Matthew S., and Michael C. Horowitz. 2012. "When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs." *The Journal of Politics* 74(2): 323–38.

Levy, Jack S., Michael K. McKoy, Paul Poast, and Geoffrey P. R. Wallace. 2015. "Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory." *American Journal of Political Science* 59(4): 988–1001.

Lin-Greenberg, Erik. 2019. "Backing up, Not Backing down: Mitigating Audience Costs through Policy Substitution." *Journal of Peace Research* 56(4): 559–74.

Lupia, Arthur, and Mathew McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* New York, NY: Cambridge University Press.

Lupton, Danielle L. 2020. *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Ithaca, NY: Cornell University Press.

Lupton, Danielle L, and Clayton Webb. 2022. "Wither Elites? The Role of Elite Credibility and Knowledge in Public Perceptions of Foreign Policy." *International Studies Quarterly* 66(3).

Markwica, Robin. 2018. *Emotional Choices: How the Logic of Affect Shapes Coercive Diplomacy*. Oxford University Press.

Mercer, Jonathan. 2012. "Audience Costs Are Toys." *Security Studies* 21(3): 398–404.

Mintz, Alex, and Karl DeRouen Jr. 2010. *Understanding Foreign Policy Decision Making*. Cambridge University Press.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62(3): 760–75.

Moon, Chungshik, and Mark Souva. 2016. "Audience Costs, Information, and Credible Commitment Problems." *Journal of Conflict Resolution* 60(3): 434–58.

Murman, Daniel L. 2015. "The Impact of Age on Cognition." *Seminars in Hearing* 36(3): 111–21.

Nelson, Stephen C. 2017. *The Currency of Confidence: How Economic Beliefs Shape the IMF's Relationship with Its Borrowers*. Cornell University Press.

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.

Payne, John W., James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. New York, NY, US: Cambridge University Press.

Potter, Philip B. K. 2007. "Does Experience Matter?: American Presidential Experience, Age, and International Conflict." *Journal of Conflict Resolution* 51(3): 351–78.

Press, Daryl G. 2005. *Calculating Credibility: How Leaders Assess Military Threats*. Ithaca, NY: Cornell University Press.

Rathbun, Brian C. 2007. "Uncertain about Uncertainty: Understanding the Multiple Meanings of a Crucial Concept in International Relations Theory." *International Studies Quarterly* 51(3): 533–57.

———. 2019. *Reasoning of State: Realists, Romantics, and Rationality in International Relations*. New York, NY: Cambridge University Press.

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): 148–81.

Saunders, Elizabeth N. 2017. "No Substitute for Experience: Presidents, Advisers, and Information in Group Decision Making." *International Organization* 71(S1): S219–47.

Schelling, Thomas C. 1966. *Arms and Influence*. New Haven, CT: Yale University Press.

Schuessler, Julian, and Markus Freitag. 2020. "Power Analysis for Conjoint Experiments." https://osf.io/preprints/socarxiv/9yuhp/ (January 18, 2023).

Simon, Herbert A. 1997. 3 *Models of Bounded Rationality: Empirically Grounded Economic Reason*. The MIT Press. http://direct.mit.edu/books/book/4286/Models-of-Bounded-RationalityEmpirically-Grounded (December 17, 2020).

Smith, Bradley C., and William Spaniel. 2019. “Militarized Disputes, Uncertainty, and Leader Tenure.” *Journal of Conflict Resolution* 63(5): 1222–52.

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.

Tetlock, Philip. 2017. *Expert Political Judgment: How Good Is It? How Can We Know? - New Edition*. New Edition. Princeton, NJ: Princeton University Press.

Tetlock, Philip, and Dan Gardner. 2016. *Superforecasting: The Art and Science of Prediction*. New York: Broadway Books.

Tomz, Michael. 2007. “Domestic Audience Costs in International Relations: An Experimental Approach.” *International Organization* 61(4): 821–40.

Tomz, Michael, and Jessica Weeks. 2013. “Public Opinion and the Democratic Peace.” *American Political Science Review* 107(4): 849–65.

Trachtenberg, Marc. 2012. “Audience Costs: An Historical Analysis.” *Security Studies* 21(1): 3–42.

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): 526–45.

Treisman, Daniel. 2004. “Rational Appeasement.” *International Organization* 58(2): 345–73.

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.

Appendix A. Conjoint Experimental Survey Questionnaire

Part 1: Introduction to Conjoint

[After consenting to participate in the study and passing a captcha-style question meant to weed out bots, respondents are shown the following introduction.]

We want to tell you more about how this survey will work.

First, we are going to describe a scenario between two generic countries — Country A and Country B. The countries are engaged in a dispute described on the next page.

Then, we are going to present you with a table containing information about a possible version of the leader of Country A. We will ask you to read carefully about this leader, and then we will ask you some questions about the possible version of the leader for Country A.

After you have completed the questions, we will present you with the EXACT SAME general scenario about the dispute between Country A and Country B. This time, however, the table containing information about the leader of Country A will be different. You will read about this new leader, then answer the same set of questions as before.

We will repeat this process a total of ten (10) times. After you have finished those tasks, we will ask you to repeat the process three (3) times, this time answering two different questions. Finally, we will ask you to answer a short questionnaire about yourself.

On the next page, we are going to run through the scenario with you. Click ">>" to proceed.

[NEW SCREEN]

Scenario: Please carefully read the following scenario between two general countries (Country A and Country B). You will be asked questions about this scenario throughout the survey -- *this scenario will always be exactly the same throughout the survey*. Y*ou will be able to reference the scenario again later in the survey.*

- Country A is involved in a dispute with its neighbor, Country B, over territory along their shared border.
- Country B claims that a small piece of territory within Country A rightfully belongs to Country B. Very few people from Country A live in the disputed territory, and the territory is not economically valuable.
- Country B has troops near the border that could be used to invade the disputed territory.
- The leader of Country A must consider their options and make a decision about how they will respond to Country B. They must choose whether to appear committed OR to concede and avoid a fight.

Did you read the scenario carefully? Please do so if you have not already, then hit "Yes" to continue.

- Yes

[NEW SCREEN]

Pre-Treatment Factual Manipulation Check: In the scenario you just read about, what was the cause of the dispute between Country A and Country B?
(Randomized Order:)

- Country B claimed that a piece of territory in Country A rightfully belonged to them
- Country B invaded Country A's territory for its natural resources
- Country B's leader insulted Country A's leader on national television
- Country B imposed trade sanctions on Country A
- Unsure/Don't Know

[Note: Respondents who do not select the correct response to the factual manipulation check will be forced to exit the survey and will be paid partial compensation for their time up to that point.]

[NEW SCREEN]

Thanks for reading the scenario carefully. In this study, the scenario you read will always be the same. But we are going to give you additional information about the leader of Country A and the situation they face in tables. Here is a sample table of what that will look like:

Leader Profile	
Age	55 years old
Time as Leader	7 years
Foreign Affairs Experience	Yes
Military Experience	No
Academic Records	Very good
More Situation Details	
Intel on Country B	Very reliable
Fatigue	Jet lagged from international travel
Stress	None.
Military Strength	Slightly stronger

In the sample table above, we provide additional information about one possible version of the leader of Country A. We provide five pieces of information in the "Leader Profile" section:

1. **Age**
2. **Time as Leader**: The length of time served as leader of Country A.
3. **Foreign Policy Experience**: Whether the leader worked in foreign affairs in a non-military capacity prior to becoming leader.
4. **Military Experience**: Whether the leader has or has not served in the military.
5. **Academic Performance**: The leader's public academic records from their time as a student.

Under the "More Situation Details" section, we also provide four pieces of information about the situation the leader finds themselves in when deciding whether to appear committed to defending the territory OR conceding the territory to avoid a fight:

1. **Intel on Country B**: The leaders have access to information about Country B from their intelligence agencies. But how reliable or accurate is the information they receive likely to be?
2. **Fatigue**: Leaders may be fatigued or tired when making decisions, perhaps due to jet lag or illness.
3. **Stress**: Leaders may have other things causing stress when making decisions, such as the illness or injury of a family member.
4. **Military Strength:** Some leaders may have access to somewhat stronger or weaker militaries compared to Country B.

[NEW SCREEN]

Consider this leader and the situation they faced when deciding whether to appear committed to defending the territory or concede the territory and avoid a fight with Country B:

Leader Profile	
Age	55 years old
Time as Leader	7 years
Foreign Affairs Experience	Yes
Military Experience	No
Academic Records	Very good
More Situation Details	
Intel on Country B	Very reliable
Fatigue	Jet lagged from international travel
Stress	None.
Military Strength	Slightly stronger

Scenario Continued: After considering their options, the leader of Country A went on public TV and committed to doing "whatever it takes" to defend the disputed territory from Country B.

Two weeks later, Country B invades the disputed territory. The leader of Country A now must decide whether to stand firm and defend the disputed territory OR back down and concede the territory to avoid a fight.

Sample DV: How likely do you think it is that the leader of Country A will now stand firm and defend the disputed territory?

- Extremely likely
- Very likely
- Somewhat likely
- Somewhat unlikely
- Very unlikely
- Extremely unlikely

Part 2: Conjoint and Outcomes x 10 times

[Note: The previous Sample DV is non-randomized and is used as practice for introducing the conjoint task. It will not be one of the ten profiles completed per respondent for the analysis.]

Now, consider a different leader and the situation they faced when deciding whether they should publicly commit to defending the territory or concede the territory and avoid a fight with Country B:

[Randomly Generated Leader Profile Here]

Scenario Continued: After considering their options, the leader of Country A went on public TV to say they would "do whatever it takes" to defend the disputed territory from Country B.

Two weeks later, Country B invades the disputed territory. The leader of Country A now must decide whether to stand firm and defend the disputed territory OR back down.

DV [Rounds 1 – 10]: How likely do you think it is that the leader of Country A will now stand firm and defend the disputed territory?

- Extremely likely
- Very likely
- Somewhat likely
- Somewhat unlikely
- Very unlikely
- Extremely unlikely

[NEW SCREEN]

[Repeat this section 10 times]

[Repeat this section an additional 3 times, replacing the DV question with the manipulation-check questions below]

We are now going to ask you two different questions about the leader of Country A and the situation that they faced when deciding whether to publicly commit to defend the territory or to concede and avoid a fight. Everything else remains the same.

Manipulation Check 1 (Ability): When it comes to making decisions about Country A's foreign policy, how skilled do you think the leader is likely to be?

- Extremely skilled
- Very skilled
- Moderately skilled
- A little bit skilled
- Not at all skilled

Manipulation Check 2 (Environment): When it comes to the situation the leader faced when deciding whether to appear committed or concede, how difficult do you think it was?

- Extremely difficult
- Very difficult
- Moderately difficult
- A little bit difficult
- Not at all difficult

Part 3: Demographics and Partisan Affiliation

[Note: These variables are post-treatment and cannot be used as covariates; however, they can be used to describe the sample's composition].

gender: Which of the following best describes your gender?

- Male
- Female
- Other [Text entry]

race: What racial or ethnic group best describes you? Please check all that apply.

- White
- Black or African-American
- Hispanic or Latino
- Asian or Asian-American
- Native American
- Middle Eastern

educ: What is the highest level of education you have completed?

- Less than a high school degree or equivalent
- High school degree or equivalent (for example: GED)
- Some college, but no degree
- 2-year college degree/Associate's degree
- 4-year college degree/Bachelor's degree
- Postgraduate degree (MA, MBA, MD, JD, PhD, etc.)

[NEW SCREEN]

mil_self: Have you ever served or are currently serving in the U.S. military?

- Yes - Previous Military Service
- Yes - Current Military Service
- No - No Previous or Current Military Service

mil_fam: Do you have any immediate family members (parents, siblings, children) who have ever served or are currently serving in the U.S. military?

- Yes - Previous Military Service

- Yes - Current Military Service
- No - No Previous or Current Military Service

[NEW SCREEN]

married: What is your marital status?
- Single
- Married
- Divorced
- Widowed

children: Do you have any children?
- Yes
- No

age: in what year were you born?
- [Text entry]

[NEW SCREEN]
pid4: Generally speaking, do you usually think of yourself as a Democrat, a Republican, or as an independent?
- Democrat
- Republican
- Independent
- Other party (please specify) [Text entry]

dstr: [Display only if pid4 == "Democrat"] Would you call yourself a strong Democrat or a not very strong Democrat?
- Strong Democrat
- Not very strong Democrat

rstr: [Display only if pid4 == "Republican"] Would you call yourself a strong Republican or a not very strong Republican?
- Strong Republican
- Not very strong Republican

lean: [Display only if pid4 == "Independent" or "Other party (please specify)"] Do you think of yourself as closer to the Republican Party or to the Democratic Party?
- Democratic party
- Republican Party
- Neither

[NEW SCREEN]

ideo: Where would you place yourself on this scale?

- Extremely liberal
- Liberal
- Slightly liberal
- Moderate; middle of the road
- Slightly conservative
- Conservative
- Extremely conservative

Part 4: Debrief

Thank you for completing the survey. The next screen will redirect you to receive compensation from MTurk for completing this survey. Before that occurs, we would like to debrief you on the study you just participated in.

We as researchers were interested in assessing which leader traits and situational challenges would influence your views of leaders' credibility during the crisis. Each scenario was randomly generated with different leader traits and situational challenges so that we could assess which additional information was important to your assessments of the leader of Country A's credibility. We hope you found the survey easy enough to understand and enjoyable!

If you have any questions/concerns about this survey, please contact... For answers to any questions you may have about your rights as a research subject, contact... Please reference protocol ID#....

Appendix B. Design Table

Question	Hypothesis	Sampling Plan	Analysis Plan	Interpretations
How do dispositional leader attributes (age, time as leader, military experience, foreign policy experience, and academic records) affect evaluations of signal credibility during crisis bargaining?	We hypothesize that being middle-aged vs. elderly, serving longer as leader, having military experience, having foreign policy experience, and having better academic records will all increase signal credibility during crisis bargaining (H1).	To test this hypothesis, we will collect 1,200 responses on MTurk via CloudResearch. All 1,200 respondents will have to pass a captcha bot check and a factual manipulation check. Each respondent will be asked to complete ten conjoint profiles so that N=12,000.	We will test H1 with a linear model that regresses evaluations of signal credibility on each leader and environmental factor (see below) with clustered standard errors at the respondent level and respondent fixed effects. We will use one-tailed tests at $p<0.05$.	If a coefficient for a given leader attribute is positive and has an associated p-value less than 0.05, we will interpret the result as offering support for H1. This result would imply that the attribute exerts a significant effect on signal credibility. If instead the associated p-value is not less than 0.05, we will not interpret that attribute as offering evidence that is consistent with H1.
How do environmental factors (intel, stress, and fatigue) affect evaluations of signal credibility during crisis bargaining?	We hypothesize having better intelligence on the other country and being less fatigued/stressed will increase signal credibility during crisis bargaining (H2).	To test this hypothesis, we will collect 1,200 responses on MTurk via CloudResearch. All 1,200 respondents will have to pass a captcha bot check and a factual manipulation check. Each respondent will be asked to complete ten conjoint profiles so that N=12,000.	We will test H2 with a linear model that regresses evaluations of signal credibility on each leader and environmental factor (see above) with clustered standard errors at the respondent level and respondent fixed effects. We will use one-tailed tests at $p<0.05$.	If a coefficient for a given environmental variable is positive and has a p-value of less than 0.05, we will interpret the result as offering support for H2. This result would imply that the attribute exerts a significant effect on signal credibility. If instead the associated p-value is not less than 0.05, we will not interpret that attribute as offering evidence that is consistent with H2.
Are the effects of leader ability conditioned by the environmental difficulty during crisis bargaining, and vice versa?	We hypothesize that the effects of increasing leader ability will be less positive when the environment is easy, or conversely, that the effects of decreasing leader ability will be the most negative when the environment is difficult (H3).	To test this hypothesis, we will collect 1,200 responses on MTurk via CloudResearch. All 1,200 respondents will have to pass a captcha bot check and a factual manipulation check. Each respondent will be asked to complete ten conjoint profiles so that N=12,000.	We will test H3 by creating additive indices of the five leader attributes and two environmental attributes, respectively, using a linear model to regress signal credibility on the indices and an interaction between them. We will use clustered standard errors, respondent fixed effects, and one-tailed tests at $p<0.05$.	If the interaction term is negative and has a p-value less than 0.05, we will interpret the result as offering support for H3. This result implies that as the environment gets easier, the effect of leader ability on credibility is less positive. If instead the associated p-valiue is not less than 0.05, we will not claim that the effects of increasing leader ability are attenuated in easier environments.

Appendix C. Minimum Detectable Effects and Statistical Power

Minimum Detectable Effect – AMCEs of Individual Components

Using the *clusterPower* R package, we find a 0.025 minimal detectable effect size for a 1,200-respondent study with each respondent providing ten observations (N=12,000, α=0.05, one-tailed tests) with 95 percent power. We adjust the calculation for the loss of statistical power due to the clustered design using the pilot data's intraclass correlation (ICC) of 0.088 and its within-cluster variance of 0.085, as well as the degrees of freedom used for the respondent fixed effects. The Cohen's *d* minimum detectable effect size with these parameters is 0.025. Although the use of respondent fixed effects reduces power by using one degree of freedom per respondent, this loss of power should be partly offset, or even exceeded, due to the gains in precision that come from partialing out heterogeneity due to respondents' differential propensities to rate leaders as more or less credible regardless of treatment effects. The minimum detectable effect sizes we provide here are thus conservative estimates.

0.025 is a very small Cohen's *d*. However, we note the traditional "large effect size (d=0.80), medium effect size (d=0.50), small effect size (d=0.20)" rule of thumb for Cohen's *d* is usually misleading when estimating AMCEs, as very small Cohen's *d* correspond to substantively large AMCEs. Schuessler and Freitag (2020), for example, find a Cohen's *d* of 0.20 corresponds to an AMCE of 10 percentage points, which is larger than three-fourths of ACMEs in their review of 15 highly-cited political science conjoints. Thus, a minimal detectable effect size of interest of 0.025 does not necessarily correspond to a trivial effect. Further, although we think leader ability and situational difficulty have large effects on evaluations of signal credibility, each individual attribute we use as indicators of leader ability or environmental difficulty may only have a small additive effect on signal credibility. Indeed, this is what we found in our pilot study; leader and situational attributes affect signal credibility, but the additive effect of any given attribute ranges from about 2- to 10-percentage points. An N=12,000 study would be well-powered for detecting small, but meaningful, effects for the individual component AMCEs.

Statistical Power – Leader Ability*Environmental Difficulty Interaction

Using the *InteractionPoweR* package (Baranger et al. 2022), we next perform power calculations for the interaction effect between the leader ability index and the environmental difficulty index. The package uses Pearson correlation coefficients as inputs and outputs. In the pilot, we estimate a *standardized* interaction effect of -0.034.[9] Given our proposed study parameters (N=12,000, α=0.05, one-tailed tests), we have 98 percent power to detect an effect of this size. We maintain 95 percent power for an interaction of -0.031. Again, these are conservative estimates of power given we are doing these power calculations while subtracting degrees of freedom for respondent fixed effects, but not accounting for the increase in precision from using respondent fixed effects.

[9] This interaction effect seems attenuated relative to that we report when using variables that take values between 0 and 1 because it takes a different (i.e., longer) range of values. An interaction of this magnitude is *not* trivially small; it corresponds to a -0.17-point interaction effect on the 0 to 1 scale in the pilot study.

Appendix D. Tables Presenting Pilot Results

Table D.1: AMCE

	OLS
Age	0.030
	(0.011)
Time in Office	0.042
	(0.010)
Foreign Policy Experience	0.045
	(0.011)
Military Experience	0.102
	(0.012)
Academics	0.028
	(0.015)
Intelligence	0.096
	(0.017)
Fatigue / Stress	0.042
	(0.015)
Num.Obs.	2494
R2	0.249
RMSE	0.21

Estimated with respondent-fixed effects and cluster-robust SE.

Table D.2: Indices

	Additive	Interactive
Leader Ability	0.253	0.350
	(0.028)	(0.070)
Decision Environment	0.130	0.201
	(0.025)	(0.055)
Ability x Environment		-0.174
		(0.113)
Num.Obs.	2494	2494
R2	0.235	0.236
RMSE	0.21	0.21

OLS estimated with respondent-fixed effects and cluster-robust SE.