Cheap Tweets?:
Crisis Signaling in the Age of Twitter

Benjamin Harris
Erik Lin-Greenberg
MIT Department of Political Science

6 April 2023

Abstract

World leaders are increasingly turning to social media to engage in crisis signaling. This raises important questions about the effects of emerging communication technologies on international politics. In particular, are threats issued via social media seen as more or less credible than those issued through traditional channels such as official government statements? Using survey experiments fielded both on a unique cross-national sample of foreign policy experts in the United States, India, and Singapore and on a U.S. public sample, we find that threat medium generally generates no significant difference in perceived credibility among members of the public and national security experts. Put differently, tweeted threats are not seen as "cheaper talk" than threats issued through more traditional channels. This project extends work on crisis signaling, elite decision-making, and the domestic politics of international relations by taking into account an increasingly common technology.

Keywords: signaling, social media, political communication, crisis bargaining, credibility

We thank Kjell Engelbrekt, Mauro Gilli, Jonas Heering, Rich Nielsen, and workshop participants at Georgetown University, the University of Oslo, the University of Pennsylvania, the Weatherhead Center for International Affairs, and ISA 2022 for helpful feedback on earlier drafts. Ella von Baeyer provided research assistance. Any remaining errors are ours alone.

As tensions between the United States and Iran surged in January 2020, President Donald Trump took to Twitter: "Let this serve as a WARNING that if Iran strikes any Americans, or American assets, we have targeted 52 Iranian sites…and those targets, and Iran itself, WILL BE HIT VERY FAST AND VERY HARD."[1] Threats like these, delivered in-real time and in full public view, have become increasingly common as policymakers around the world turn to Twitter and other social media outlets, raising important questions about the effects of social media on international politics. In particular, how should scholars and policymakers think about the credibility of interstate signals in an era where social media use is ubiquitous? Do national security practitioners perceive threats issued via social media as equally credible as those issued through traditional channels such as official government statements? Do domestic audiences in the threat-sending state view their leader's Tweet-based threats as credible? Does the channel through which a threat is issued affect perceived credibility among the threat-receiving state's population?

The credibility of a leader's statements—particularly during crises—is central to the study of international relations. Threats or promises that are perceived as credible can shape another state's behavior, while those that are viewed as "cheap talk" often have little sway on a rival's behavior, degrade the reputation of the sending actor, and generate audience costs among the sending state's domestic population. As such, leaders must make their signals credible to multiple audiences, including members of the domestic public and foreign policy decisionmakers at home and abroad. When sending these public signals, leaders and governments have historically relied on public addresses or written public statements.

The proliferation of social media platforms as a low-cost means of rapidly communicating information to multiple audiences has increasingly led leaders and governments to use these

[1] The Tweet was widely covered in media. See, for example (Diamond, Kelly, and Clary 2020)

channels as diplomatic messaging tools (Seib 2012; Collins, DeWitt, and LeFebvre 2019; Ven Bruusgaard and Kerr 2020). Indeed, at least 189 world leaders maintain Twitter accounts and routinely tweet about policy matters—including on foreign policy issues (*Twiplomacy Study 2020*). Studying whether and how elite social media use affects crisis signaling therefore has important implications for international relations scholarship. Specifically, it allows us to explore how well core IR concepts related to interstate signaling hold up to changing trends in technology use.[2]

To study whether and how a leader's use of social media can affect perceptions of crisis signaling, we first synthesize disparate literatures from international relations, communications studies, and political psychology to lay out three potential outcomes for how a leader's use of social media to issue threats affects perceptions of threat credibility. The domestic public in threat-issuing and threat-receiving states and national security decisionmakers at home and abroad might view a tweet-issued threat as less, equally as, or more credible than a threat issued using a more traditional platform.

To empirically assess the effect of threat medium, we leverage data from a series of complementary survey experiments fielded on U.S. public samples and a cross-national sample of foreign policy practitioners and experts in the U.S., India, and Singapore. The experiments vary whether the U.S. president uses a tweet or official White House statement to issue a threat during an interstate crisis. This multi-pronged research design allows us to assess whether and how a leader's use of social media to issue threats affects perceptions of threat credibility among two core audiences: the domestic public and international policy practitioners.

[2] On the importance of studying "digital diplomacy," see (Hedling and Bremberg 2021).

Our experiments reveal that, in general, varying threat medium between Twitter and more traditional channels generates little difference in perceived threat credibility. Put differently, threats issued via social media are not necessarily seen as "cheaper talk" than more traditionally-issued threats. At most, a president's threat issued using Twitter is seen as very slightly less credible among the U.S. general public than an identical threat issued via official White House statements. Among national security experts, however, tweets are viewed as being just as credible as official White House statements. In a series of follow-on experiments, we explore the effects of varying leaders, tweet wording, and the threat-issuing state.

We seek to contribute to two key areas of international relations scholarship. First, this project extends work on crisis signaling by taking into account new technologies that leaders use to transmit signals. Existing explanations for signal credibility often invoke a leader's past behavior, her disposition, the degree to which she has taken steps that sink costs or tie hands, and her ability to follow through on threats or promises (Fearon 1997; Kertzer 2016; McManus 2017; Yarhi-Milo, Kertzer, and Renshon 2018; Lupton 2020). These studies, however, typically overlook whether the channel through which a leader issues a threat affects its credibility.[3] While the academic literature has not examined whether communication medium affects threat credibility, policy pieces and journalistic accounts tend to assume that tweets are less credible, potentially leading to a premature consensus on an under-researched topic (Shapiro 2018; Ghitis 2019; Shepp 2018).

Second, the project has implications for scholarship on the politics of emerging technologies, particularly social media. Recent work on social media highlights how platforms like

[3] Past work has compared private signaling to public signaling (For instance, Carson and Yarhi-Milo 2017). We focus here on assessing whether variation in public platform (i.e., social media versus press releases) affects perceived credibility.

Twitter add to the chorus of voices the public must sort through when developing positions on foreign policy issues (Baum and Potter 2019; Kreps 2020), explores specific cases (Duncombe 2017; Williams and Drew 2020), studies the relationship between social media use and escalation (Narang and Williams 2022), assesses domestic political implications (Mir, Mitts, and Staniland 2022), and highlights the dangers of misinformation (Chesney and Citron 2018). This paper builds upon this growing body of research by more fully exploring social media's effects on issues on the core IR concepts of crisis signaling and perceptions.

THEORY

Crisis signaling is a prominent feature in international relations scholarship. Whether a leader's threats and promises are viewed as credible or as "cheap talk" has important implications for both international and domestic politics. On the international stage, signaling is critical to deterrence, compellence, and broader diplomatic efforts. A signal that other states perceive as truthful can shape a rival's behavior without the need for further escalation. Credibility is therefore a key determinant of which deterrent or compellent signals influence an adversary's actions and which are ignored. Drawing from Schelling's lexicon, a credible signal that convincingly demonstrates the "power to hurt" can coerce without resorting to potentially more costly and dangerous "brute force" tactics (Schelling 1966). Domestically, leaders who engage in cheap talk may face political punishment from domestic audiences.[4] This can diminish a leader's ability to pursue her policy agenda (Gelpi and Grieco 2015), raise questions about her competence (Smith 1998), and generally decrease public support for the leader (Tomz 2007).

[4] The concept of audience costs is the subject of an ongoing scholarly debate. For instance, see (Snyder and Borghard 2011; Trachtenberg 2012)

Simply put, credibility is the perceived likelihood that an actor will follow through on her threats or promises. Prominent explanations suggest that leaders can enhance signal credibility by "sinking costs" through military deployments that are *ex ante* costly or by "tying hands" and making pledges that are politically or reputationally costly to back down from *ex post* (Schelling 1967; Fearon 1997). A leader's willingness to accept these potential costs is thought to help observers distinguish credible from non-credible threats.[5]

Scholars have long debated the underlying determinants of credibility (Jervis, Yarhi-Milo, and Casler 2021). Some work suggests signal credibility is shaped primarily by a leader's reputation—whether they followed through on past threats and promises (Guisinger and Smith 2002; Lupton 2020). Other scholars suggest that reputation plays a less important role than an actor's capability to follow through on a threat or promise, specifically whether she possesses the military or political wherewithal to do so (Press 2007; McManus 2017). Scholars have also explored the extent to which a leader's individual-level dispositions affects how they perceive a rival's threats (Yarhi-Milo, Kertzer, and Renshon 2018).

Less well studied is the extent to which the channel or medium through which a signal is transmitted affects perceptions of the signal's credibility. Existing IR research that explores variation in signal transmission typically examines whether publicly visible signals are perceived differently than secret signals visible only to decisionmakers (Kurizaki 2007; Carson and Yarhi-Milo 2017). The traditional logic suggests leaders tie their hands through publicly issued threats and promises, and only make statements they intend to follow through on to avoid domestic political repercussions for backing down (Fearon 1994). More recent scholarship, however, suggests that leaders may perceive covert threats that are not publicly visible as highly credible

[5] To be clear, this logic is up for debate. For instance, Fuhrmann and Sechser (2014) find no evidence of the sinking costs logic.

(Carson and Yarhi-Milo 2017). Paying greater attention to threat medium is important as communications scholars have long suggested that information-transmission channels can shape how signals are perceived. Indeed, as Marshall McLuhan (1964) famously argued, "the medium is the message."

Signaling in an Era of Social Media

Although existing IR scholarship has studied the distinction between public and secret signals, it has largely overlooked whether variation in the public channels that leaders use to issue threats affects perceived signal credibility. This absence is problematic given that leaders and governments now select from a growing menu of public platforms for use in crisis signaling, including a variety of social media outlets. While debates over the reputational, dispositional, and capabilities-based determinants of credibility remain unresolved, we bracket these factors and focus on assessing whether public signaling channel affects perceived credibility.

New communications technologies are transforming how leaders engage in coercive diplomacy and crisis bargaining. Social media platforms, which now feature prominently in political communication at the domestic and international levels, allow users to rapidly create and disseminate information about important political events without the need to wait for gatekeepers or intermediaries—such as traditional media outlets—to broadcast reports (Zeitzoff 2017; Kreps 2020).[6] Moreover, increased global use of platforms like Twitter, Facebook, and Instagram enables leaders to quickly convey information to massive international and domestic audiences with relatively little cost. Indeed, in 2021, Twitter had more than 206 million daily users while Facebook had 2.8 billion users active on a monthly basis (Statista 2021).

[6] To be sure, traditional media outlets often rebroadcast social media posts. Rebroadcasts, however, still typically identify the original source.

World leaders have recognized social media as an inexpensive and instantaneous, yet highly visible channel for communicating with constituents, rivals, and allies. Although the use of platforms like Twitter is commonly associated with Donald Trump's contentious and saber-rattling messaging, Barberá and Zeitzoff (2018) find that more than three-quarters of world leaders are active on social media. Beyond leaders themselves, government agencies and other officials maintain social media accounts and regularly use Twitter and other similar channels to communicate with domestic and international audiences, including during high-stakes international crises (Williams and Drew 2020). In March 2018, for instance, the Russian Embassy in London tweeted, "Any threat to take 'punitive' measures against Russia will meet with a response. The British side should be aware of that." This warning came on the heels of the British government accusing Russia of poisoning a former Russian spy residing in the United Kingdom.[7]

These social media-issued statements fall along a continuum of specificity. Some, like Trump's January 2020 threat to strike Iranian targets, are explicit and include clear red lines. Others are less direct. For instance, in January 2018, Trump warned North Korean leader Kim Jong Un that "I too have a Nuclear Button, but it is a much bigger & more powerful one than his, and my Button works!"[8] These examples highlight the role social media platforms, particularly Twitter, play in contemporary crisis signaling.

To be sure, social media operates in parallel with other public and private channels for diplomatic communication. A leader might, for instance, issue a threat via Twitter and through an official statement, while simultaneously sending private diplomatic communiques to a target state. While the interaction of signals in different mediums will affect threat credibility, exploring the effect of social media in isolation is still important. If public audiences are not privy to the private

[7] For a detailed summary of the tweets, see (Williams and Drew 2020).
[8] @realDoanldTrump, 2 January 2018, https://twitter.com/realDonaldTrump/status/948355557022420992.

signals that leaders send to their rivals, the public may subsequently base assessments only on publicly visible signals. Moreover, the emerging popular consensus that tweets are a less serious and less credible communication medium suggests a need to explore whether the intrinsic quality of social media shapes audience perceptions.

Threat Credibility

How might a leader's use of social media for coercive diplomacy affect perceived signal credibility? Research from psychology and communication studies yield mixed predictions. Some studies suggests that information conveyed on outlets like Twitter is perceived as less credible than information presented on more traditional sources (Johnson and Kaye 2014), while other projects find that social media sources are considered just as credible.[9] Other scholarship examines how characteristics of social media posts—such as formatting, originating source, or metrics such as the number of "likes"—affect perceived credibility (Morris et al. 2012; Westerman, Spence, and Van Der Heide 2012; Jahng and Littau 2016). These studies, however, are narrowly focused on what elements of a social media post bolster credibility, rather than exploring how a leader's choice of transmission medium affects the credibility of a specific message. More importantly for IR scholars, these studies typically focus on domestic political issues rather than questions related to foreign policy or international security.

The first possibility is that social media-issued crisis signals might be seen as lacking credibility vis-à-vis the same signal issued using more traditional platforms. Tweets in particular could be perceived as "cheap talk" since a leader can issue a tweet without the extensive interagency coordination required to craft and publish a more official statement.[10] Former U.S.

[9] This work focuses on comparing social media with traditional media outlets rather than with government outlets.
[10] Of course, Tweets can also go through a similar staffing process, but the ability of a leader to issue a Tweet on his or her own means they can circumvent the staffing process.

Secretary of Defense Mark Esper, for instance, explained that President Trump often developed his Twitter messages "with just one or two other people…They were never coordinated outside of the Oval Office, it seemed, let alone with the departments (Esper 2022, 158)." Trump's threat to strike 52 Iranian targets was tweeted, "without any consultation with military and civilian leaders responsible for effecting such a mission (Esper 2022, 157)" and the military was unprepared to execute such an operation that, according to Secretary Esper, "clearly had no operational logic behind it (Esper 2022, 156)."

Indeed, leaders often issue dozens of tweets a day, potentially making these 280-character messages appear unimportant or insignificant. Moreover, a tweet may be perceived as less coordinated than more formal statements—like official press releases—in which members of the interagency typically work together to produce a cohesive and intelligible message. Interagency coordination should prevent leaders from making baseless statements because advisors and policymakers understand that issuing empty threats and promises can generate political and reputational consequences. As a result, policymakers may attempt to block dissemination of non-credible messages. For these reasons, members of the public and international leaders might not view a leader's tweets as an indication of an administration's actual policies.[11]

The second possibility is that domestic and international audiences may perceive threats issued via social media to be just as credible as those issued using traditional channels. Making a threat on social media is an inherently public act that potentially ties a leader's hands in the same way as a public threat that is conveyed using more traditional channels. A tweet, for instance, can reach millions of Twitter users worldwide, even without being magnified by coverage in traditional media outlets. Indeed, a leader's tweet might disseminate far faster than if the message were issued

[11] These concerns can be magnified when a specific leader is known to make inaccurate statements or when a leader's tweets do not align with statements of other government actors or institutions.

through a press release or a formal statement to traditional media. Given the highly public nature of many social media platforms, a tweet provides a written record of a leader's threats or promises that the public can use to hold a leader accountable should she subsequently fail to follow through.

According to the "tying hands" logic of costly signaling, tweets should be perceived as credible. Since issuing an "empty tweet" is highly visible and could lead to reputational and political consequences, a leader should only tweet a threat she intends to carry out. Moreover, members of the public and international decision-makers may not even differentiate between a tweet-issued threat and a public threat issued through a more traditional channel. Put differently, they might assume tweets and formal messages go through the same drafting process (or that staffers issue a leader's tweets) and are therefore one and the same.

The third possibility is that a leader's tweeted threat might be perceived as more credible than the same threat disseminated through more traditional channels. This should be the case if foreign decision-makers and members of the public believe social media posts represent the unfiltered thoughts of the tweeting leader (and also think the leader is able to implement her desired policies). Underlying this line of reasoning is the belief that the most credible signals are those that feature unplanned, off-the-cuff comments (Jervis 1970). These statements, which often lack coordination with other government entities and are typically imbued with emotion, are thought to offer valuable insight into a leader's thinking.

We believe it is unlikely that a tweet-issued threat will be viewed as more credible than a threat issued via a more traditional channel. A formal statement should carry more weight than a leader's social media musings given that the former likely requires greater whole of government coordination to develop and issue, and should therefore be viewed as representing the state's true intent. This leads to our first testable hypothesis:

H_1: National security experts and members of the general public are likely to perceive a threat issued via social media as less credible than a threat issued via an official statement.

In summary, there are three potential outcomes: the hypothesized (and preregistered) outcome in which tweet-issued threats are perceived as less credible; the null hypothesis—that the public views tweets as equally credible as official statements; and the inverse of H_1—that the public finds tweets more credible than official statements.

METHOD

To assess whether and how social media use as a signaling medium affects perceived threat credibility, we turn to original survey experiments that manipulate whether the U.S. president issues a threat to a rival using Twitter or an official written White House statement. To study the effects on both domestic public and international expert audiences, we adopt a two-prong recruitment strategy that includes complementary experiments fielded on a U.S. public sample and on a cross-national sample of national security practitioners and experts in the U.S., India, and Singapore.[12] While past studies have examined specific cases in which leaders used Twitter for crisis signaling (Duncombe 2017; Williams and Drew 2020), our experimental approach allows us to precisely identify whether the use of Twitter for signaling affects threat credibility.

Our survey instrument presents respondents with a hypothetical, but plausible, crisis that involving the United States and Iran. All respondents are told:

> "Over the past several months, the Iranian government has provided funding, training, and weapons to militia groups that have launched several attacks on U.S. forces and partners throughout the Middle East. Earlier this week, Iranian-backed militias attacked two oil tankers in the Red Sea that were transporting fuel to the United States and fired rockets at the U.S. Embassy in Yemen. The attacks caused significant damage to the oil tankers and the embassy and killed eight people, including one American."

[12] For more on the utility of "paired experiments" and "complementary designs" that feature experiments fielded on both mass and elite samples, see (Kertzer and Renshon 2022).

We then randomly assign respondents to one of two experimental conditions that manipulate whether "President Biden made his first statement about the situation by issuing the [Tweet *or* official White House Press Release] below." We describe this as the "first statement" to suggest that the president has not yet made similar announcements on other platforms.

Unlike many survey experiments that present respondents solely with textual information, we attempt to enhance the realism of the treatments by presenting realistic facsimiles that closely mirror the actual format of Tweets and White House press releases (Figures 1a and 1b). While the experiment varies the medium through which President Biden's threat is issued, the treatments hold threat content constant. We then ask respondents a series of questions about their perceptions of the threat's credibility. In follow-on experiments, described below, we vary the identity of the president (i.e., Trump or Biden) and the wording of the tweet, and present a scenario in which Iranian officials are issuing threats against the United States.

Figure 1a: Tweet Treatment

Figure 1b: Press Release Treatment

We made several key design choices when developing the main experiment. First, we explicitly name President Biden to avoid introducing ambiguity about the president's identity. This is critical because foreign policy elites and members of the public are thought to consider the leader's reputation when judging the threat's credibility (Lupton 2020). Leaving the president unnamed could lead respondents to make assumptions about the president's identity, potentially

leading to assessments based on assumptions rather than on the treatment of interest. Moreover, recent work finds that "real and highly salient cuegivers" often generate stronger effects than fake ones (Brutger et al. 2021). Second, we name Iran as the target of President Biden's threats. As with identifying the president, naming a specific rival mitigates the risk that respondents will make assumptions about the identity or other attributes of the unnamed rival (Dafoe, Zhang, and Caughey 2018). Finally, we use Twitter as our social media treatment. To be sure, world leaders maintain a presence on other platforms such as Facebook and Instagram, but there has been widespread government use of Twitter to issue coercive threats. We compare the tweet to a written White House press release, rather than a more dissimilar medium, like a presidential press conference. In sum, naming the president and rival country and using likenesses of Tweets or press releases allow respondents to engage with a realistic vignette.

Of course, these choices limit the generalizations that can be drawn from experimental findings. Critics might suggest several additional limitations to our research design. First, one might argue that leaders use social media to issue threats only in circumstances involving lower stakes where credibility might be less important than a high-stakes crisis. The empirical record, however, suggests that governments routinely use Twitter even during major crises involving highly contentious disputes with peer competitors. Second, critics might argue that the experimental manipulations are more salient in the experimental context than they would be in the real-world. During an actual crisis, members of the public and national security experts would receive information not only via Twitter or White House press release, but from a variety of other traditional and social media sources. Moreover, a public threat might be coupled with other signals, such as the posturing of military forces and a diplomatic communique. While we concede this is true, we believe our research design maximizes internal validity while still offering externally valid

insights. In an actual crisis, members of the public and national security experts may not be aware of the full range of actions a threat-issuing state has taken. As a result, they may formulate their judgments of threat credibility based solely on publicly available information—like tweets and official statements.

More broadly, scholars continue to debate whether findings from survey experiments are valid beyond the controlled confines of survey instruments. To be sure, experimental subjects face different stakes than those faced during actual crises and also receive different degrees of information than their counterparts confronting real-world crises. We believe, however, that survey experiments remain useful tools for studying expert and public preferences because experiments are grounded in the assumption that respondents rely on the same cognitive processes they would use when making judgments in the real world (Schelling 1961, 55).

We fielded the main experiment on a 977 respondent U.S. public sample recruited using Lucid, an online sampling service, in July 2021. Lucid relies on quota sampling to recruit samples that align more closely with U.S. Census demographics than many online convenience samples (Coppock and McClellan 2019). Lucid samples, however, are not nationally representative across all dimensions. Our sample, for instance, skews older than a national sample; underrepresents Hispanic Americans and those with household incomes greater than $100,000; and overrepresents college educated Americans and white Americans. Still, Lucid samples are generally more representative than other online convenience samples (e.g., Amazon MechanicalTurk).

To recruit our cross-national foreign policy expert sample, we follow Clark (2021) and use LinkedIn messages targeted at government and think tank employees who work for institutions that make or advise on national security policy (e.g., Ministry of Defense, Ministry of Foreign

Affairs).[13] These respondents' domain-specific expertise and experience offers unique analytic leverage for understanding the perceptions of foreign policy decisionmakers (Dietrich, Hardt, and Swedlund 2021; Kertzer and Renshon 2022). We field the expert survey in the United States, India, and Singapore because these states feature professional, English-speaking, foreign policy bureaucracies and vibrant think tank communities that engage in policy debates.

Results from U.S. expert respondents allow us to probe whether domestic experts in the threat-issuing country perceive tweet-issued threats differently than those issued through more traditional channels, while responses from Singapore and India shed light on how international audiences perceive these threats.[14] Expert participants completed the survey between August and October 2021.[15] Given our online recruitment strategy, we assume both our public and expert respondents regularly use the internet.

MAIN EXPERIMENT RESULTS

The main experiment suggests that signaling medium has at most, a small effect on perceived threat credibility. In a simple model with no demographic covariates, variation in communication medium has no significant effect on perceived credibility among both the public and elites.[16] When standard demographic controls (e.g., gender, political ideology, race, etc.) are included, a Twitter-issued threat is viewed as slightly less credible among the public sample than

[13] See Appendix A for additional information on the recruitment strategy. The appendix includes the recruitment text and the organizations from which expert respondents were recruited.
[14] In the ideal case, we would have also fielded an expert survey in the target country of Iran, but this was not logistically feasible.
[15] The U.S. withdrawal from Afghanistan and the Taliban's subsequent takeover occurred shortly after we began fielding our expert experiment in the U.S. (the India and Singapore surveys were fielded entirely after Kabul fell). These events triggered significant policy debates about President Biden's reputation and credibility. Because these real-world events might have influenced respondent perceptions of threat credibility, we include analysis in Appendix B that assesses whether respondents who completed the experiment before Kabul fell hold different perceptions than those who completed the survey after the Kabul's fall. We find no difference in perceptions.
[16] For the public sample, the coefficient of the tweet medium on credibility is negative. In the expert sample, it is positive, but neither meet standard thresholds of statistical significance.

a threat issued via more traditional channels (marginally significant at the $p > 0.10$ threshold). However, variation in communication continues to yield no significant results in the elite sample.

Table 1: Summary of Results from Main Experiment

Dependent Variable	Demographic Controls	Sample	Hypothesized Direction	Tweet Coefficient Direction	Statistical Significance
Credibility	No	Elite	Negative (i.e., Tweet is less credible)	Positive	No
Credibility	Yes	Elite	Negative	Negative	No
Credibility	No	Public	Negative	Negative	No
Credibility	Yes	Public	Negative	Negative	$p < 0.10$

Assessing Credibility

To assess whether the channel the leader uses to issue a threat affects perceptions of credibility, we ask both the public and expert respondents, "In your opinion, how likely or unlikely is it that the president will follow through on his threat?" We use the perceived likelihood of following through as a proxy for credibility. If a threat is perceived as credible, respondents should be more likely to believe the president will carry out the threat. Respondents rate the likelihood of following through using a 5-point Likert scale ranging from "Extremely unlikely" (1) to "Extremely likely" (5).

Among respondents in the expert sample, we find that threat medium has no statistically significant effect on the perceived credibility of President Biden's threats. Put differently, national security experts from all three countries viewed the tweet-issued threat to be just as credible as the same threat issued via a White House statement. On average, respondents across the full sample of experts viewed the credibility of tweeted threats to be 3.11 on the five-point scale compared to 3.10 for threats announced through an official statement (Table 2). Figure 2 plots the average treatment effect of threat medium on credibility with the baseline condition as a threat issued via

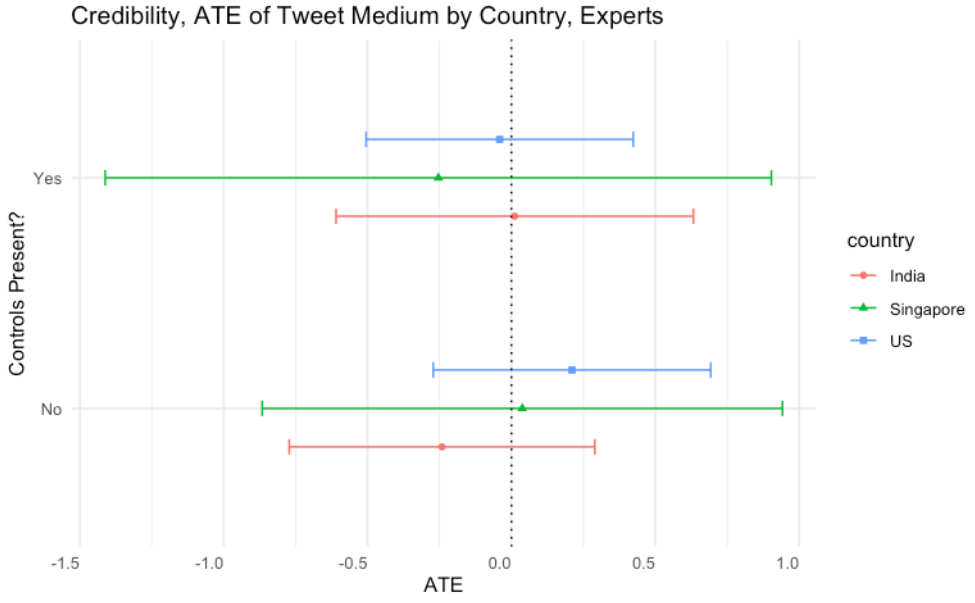
press release. The plot on the upper half of the figure accounts for several covariates including respondent age, education, gender, veteran status, and past government work experience, while the plot on the lower half does not. Although we are hesitant to draw too many conclusions due to the small sample sizes, the findings reveal no significant cross-national variation in perceived credibility. Among the demographic covariates, three results were significant at the $p < 0.05$ threshold in the full expert sample: veteran status and female identity were associated with decreased perceptions of credibility while age varied directly with perceptions of credibility, regardless of threat medium.[17]

Table 2: Mean Perceived Credibility, Expert Sample

	Full Sample	United States	India	Singapore
Tweet Mean Credibility	3.11 (0.12) *n=119*	3.21 (0.18) *n=57*	2.92 (0.18) *n=51*	3.45 (0.33) *n=11*
Press Release Mean Credibility	3.10 (0.12) *n=115*	3.00 (0.17) *n=60*	3.16 (0.20) *n=43*	3.42 (0.32) *n=12*

Standard errors in parentheses.

Figure 2: Average Treatment Effect of Tweet Medium on Credibility, Expert Sample



[17] Appendix C includes regression models featuring the full set of covariates.

We find similar results among members of the U.S. public when not including demographic controls. Respondents, on average, rated tweeted threats to be 3.22 on the five-point scale credibility compared to 3.30 for threats announced through a press release (Table 3). Although the results are not statistically significant, they are in the hypothesized direction with tweet threats, on average, perceived as slightly less credible than one issued via an official statement. When demographic controls are included, the coefficient for the Tweet medium remains negative and is statistically significant at the $p < 0.10$ threshold.

Unsurprisingly, liberal respondents are, on average, likely to view President Biden's threat as more credible than Conservatives do, regardless of threat medium. Further, older Americans are, on average, slightly less likely to view the president's threat as credible while veterans are likely to view the threat more credibly, irrespective of whether the threat is tweeted or issued via an official statement. We find no interaction effects between any of our significant demographic covariates and the Tweet medium.

Table 3: Mean Perceived Credibility, U.S. Public Sample

	U.S Public Sample
Tweet Mean Credibility	3.22 (0.06) *n=491*
Press Release Mean Credibility	3.30 (0.06) *n=486*

Standard errors in parentheses.

To assess what underpins these perceptions, we collected qualitative data collected by asking respondents to "write a sentence or two explaining [their] response." Just 2.2-percent of respondents explicitly mentioned the medium, preventing us from identifying systematic trends. Yet the limited responses yield some insight into respondents' thinking. Some responses indicate that tweets are viewed like other forms of public official statements. One respondent in the public

sample notes that "President Biden didn't write that tweet, he has people that do that for him. He would only do what a tweet says if his staff told him to." Another respondent alluded to the tying hands logic, by explaining that, "if he tweeted that he is likely to follow up on it or else people would take him as a joke." Other respondents, however, explicitly raised questions about the credibility of tweets. One respondent explained that, "A threat on social media is a weak response and indicates weak follow through." Another commented that a "threat vocalized through a tweet adds a certain level of ingenuity." Several others suggested Twitter was not the venue for engaging in crisis bargaining. "Very inappropriate if a president would do that," noted one respondent. "A tweet is not the place! Address the Nation on prime-time saying that and it would be extremely likely [i.e., credible,]" said another. In short, while some respondents explicitly described how the threat medium affected their perceptions of credibility, a vast majority did not, limiting the inferences we can make.

To further assess our results, we turn to data from a series of manipulation check questions built into the experiment fielded on the public sample. One question asked respondents to identify the treatment they received: whether they read a tweet or an official statement. Three other questions asked respondents to recall details of the scenario. Respondents who were presented a tweet were much more likely to fail the treatment manipulation check than respondents presented with an official statement.[18] In other words, respondents in the tweet treatment were more likely to incorrectly report that they had read an official White House statement (51.94%) than respondents in the official statement treatment were to incorrectly state that they had read a tweet (30.45%).[19] Despite having a higher likelihood of inaccurately identifying the treatment they received, respondents in the tweet treatment were no more likely than respondents in the official

[18] This finding is significant at the $p < 0.01$ threshold.
[19] See Appendix B for analysis of manipulation checks.

statement treatment to incorrectly recall specific vignette details including the target of the attack, the country supporting the attack, or the sea in which the oil tankers were located. These findings suggest many members of the public view tweets as official White House communications, which potentially explains the relatively small and statistically insignificant results.

In sum, our complementary public and expert experiments do not yield strong support for our hypothesis that tweeted threats should be perceived as less credible than those issued using more traditional mediums. This finding, however, offers valuable insights for the study of crisis bargaining and the politics of emerging technologies. Critically, our findings suggest threats issued via social media should not be discounted as being less credible than those issued using more conventional channels such as official statements. Moreover, given that perceived credibility hovers around a mean of three on a 5-point scale regardless of medium, the results suggest that the public generally views publicly-issued threats as only moderately credible.

Although social media-issued threats very slightly reduce perceived threat credibility for some members of the public, many respondents see little difference between a president's tweets and their official written statements. Put differently, for many members of the public, a president's tweets are themselves viewed as official written statements. Indeed, some of the qualitative responses suggest respondents believe presidential tweets are developed in much the same way as official statements—through a formal staff process—and do not represent a president's impromptu thoughts. Moreover, national security experts, on average, perceive a tweeted threat as being just as credible as one issued via an official White House statement. This is perhaps because individuals with national security experience understand that a leader's social media posts often go through a similar vetting process as official statements.

EXPLORING VARIATION IN LEADER AND TWEET WORDING

Because different leaders may use social media in different ways, we field a follow-on experiment that varies the president's identity and the level of formality in the Tweet language.[20] The follow-on experiment is similar in design to the main experiment, featuring the same crisis scenario and the president threatening retaliation to future attacks either via Tweet or official White House Statement. The follow-on, however, introduces new manipulations that allow for additional analysis. To assess whether the Tweet-issuing leader's identity has an effect on perceived credibility, we vary whether Donald Trump or Joe Biden is president. To make this plausible, we situate the scenario in September 2026, when either individual could be president. We also vary whether the text in the Tweet is formal or informal to assess how the specific language a leader uses might affect outcomes. The formal language is identical to the original experiment, while the informal language is depicted in figure 3.

Figure 3. Informal Trump Tweet

Based on the main experiment findings, we expect variation between a formal tweet and a White House statement to yield no significant differences in credibility. Informally worded tweets, however, might be viewed as less credible than formally worded ones, although any difference is

[20] We field the experiment in February 2022 on a 1383-respondent Lucid sample.

likely to be small in magnitude. When varying the president's identity, we expect heterogenous treatment effects based on respondents' party affiliation. Conditional on threat medium, Democrats are likely to view President Biden's threat as more credible than President Trump's threat.

Follow-On Results

The follow-on experiment yields results that align with those from our original experiment: neither the communication medium nor text formality exerts significant influence over perceived threat credibility. Table 4 below reports the mean perceived credibility levels for each treatment group (on a five-point scale).[21]

In a simple model without demographic controls or interaction terms, the only significant result was the president's identity: respondents rated threats issued by President Trump as more credible, regardless of threat medium. For our demographic covariates, Black and Native Hawaiian or Other Pacific Islander racial identity as well as age are correlated with an increase in perceived credibility.[22] When adding interaction terms, we find no significant effects for the interactions between the president's identity and language formality or president's identity and the Tweet medium. Unsurprisingly, we again find that liberal respondents view a threat issued by President Biden to be more credible while more conservative respondents find a threat issued by President Trump to be more credible. In sum, threat medium continues to have little effect on perceived threat credibility.

Table 4: Mean Perceived Credibility, First Follow-On Experiment

	Biden	**Trump**
White House Press Release	3.00 (0.09) *n=230*	3.39 (0.09) *n=236*
Formal Tweet	3.01 (0.09) *n=227*	3.31 (0.09) *n=232*

[21] The results of OLS regressions are contained in Appendix D.
[22] Our sample includes very few NH or OPI respondents, making this unreliable.

Informal Tweet	3.04 (0.09) *n=225*	3.10 (0.09) *n=233*

Standard errors in parentheses.

When examining the manipulation check questions, we once again find that respondents are more likely to incorrectly identify their treatment if they read a Tweet (42.09%) rather than an official statement (30.04%).[23] Oddly, respondents in the Tweet treatment are also slightly more likely to *correctly* identify the president (92.80%) compared to those in the official statement treatment (89.06%).[24]

EXPLORING VARIATION IN THREAT-ISSUING STATE

While the experiments featuring U.S. presidents shed light on perceived credibility among the U.S. public and experts in the U.S., India, and Singapore, crisis signaling is ultimately intended to shape a targeted state's behavior. To explore whether communication medium shapes how a rival's threats are perceived among a target state audience, we field a second follow-on experiment in which Iran's Supreme Leader threatens the United States. The follow-on was fielded on a public sample of 1,477 adults in the United States recruited using Lucid in January 2023.

All respondents are told:

> "After an Iranian-backed rocket attack on the U.S. Embassy in Yemen killed eight people, including one American, President Biden publicly warned the Iranian government that the United States would conduct military strikes on Iranian military facilities in Yemen if Iran supported further attacks on Americans. The Iranian Supreme Leader responded to President Biden's warning by issuing the [official statement *or* Tweet] below."

[23] This result is significant at the $p < 0.01$ threshold. See Appendix D.
[24] This result is significant at the $p < 0.05$ threshold. See Appendix D.

Figure 4a: Tweet Treatment

Figure 4b: Press Release Treatment

After respondents are presented with either the Tweet (Figure 4a) or official press release from the Office of the Supreme Leader (Figure 4b), the survey instrument asks about perceived credibility. As with the other experiments, we measure perceived credibility by asking, "In your opinion, how likely or unlikely is it that the Iranian Supreme Leader will follow through on his threat if the United States strikes Iranian facilities?" In line with the experiments described above, threat medium has no significant effect on perceived credibility of the Supreme Leader's threat. Although a tweeted threat yields a very slightly lower mean credibility level than one issued via press release, the difference is small and not statistically significant. We find that Female respondents are more likely to view the threat as credible while those with Native Hawaiian or Other Pacific Islander (NH/OPI) racial identity and those with a high school education are less likely to view the threat as credible. Because our sample includes very few NH/OPI respondents and very few respondents without a high school diploma we refrain from making inferences from these findings.[25]

Table 5: Mean Perceived Credibility, U.S. Public Sample, Iran Follow On

	U.S Public Sample
Tweet Mean Credibility	3.55 (0.04) *n=745*
Press Release Mean Credibility	3.58 (0.04) *n=735*

Standard errors in parentheses.

[25] See Appendix E for full OLS results.

As before, we find that respondents in the Tweet treatment are more likely to miss the treatment manipulation check question (26.98%) than those in the official statement treatment (14.29%).[26]

CONCLUSION

Much existing work on crisis signaling and credibility centers on a leader's reputation, a state's military capabilities, or whether a leader has taken actions that tie her hands or sink costs. These studies are generally agnostic to the medium through which a public threat is issued. Yet, leaders today have an increasing menu of platforms through which to publicly threaten rivals. Our analysis explores whether variation in the platform used to communicate coercive signals affects perceived threat credibility among both domestic and international audiences. Drawing data from a series of survey experiments fielded on a unique cross-national sample of foreign policy experts and on members of the U.S. public, we find that variation in threat channel has at most a limited effect on perceptions of threat credibility. The public sometimes views tweet-issued threats as slightly less credible than those issued using official statements, but in general, tweeted threats are seen as no different than those transmitted using more traditional channels.

Our analysis moves beyond existing scholarship on interstate signaling by taking into account increasing social media use by world leaders as a medium of coercive diplomacy. In doing so, we contribute to the study of signaling and perceptions, the domestic politics of foreign affairs, and the implications of emerging technologies on international security.

Our research raises questions about the emerging popular consensus among journalists and foreign policy analysts that Twitter is a less credible communication form than more traditional

[26] This result is significant at the $p < 0.01$ threshold. See Appendix E.

channels. Our experiments reveal that the intrinsic characteristics of Twitter itself need not decrease the perceived credibility of messages conveyed via tweet.

Our findings also point to several additional avenues for future research. Since signaling does not occur in a vacuum, scholars might study whether variation in the framing of tweeted threats affects perceived credibility. For instance, future work might explore what happens when a senior official's tweets diverge from official government statements. Or, additional experiments might explore whether the amplification of tweets through retweeting or rebroadcasting in traditional outlets shapes how people perceive the threats.

Future studies might also study other communication mediums. For instance, the effect of threat a traditional communication medium might have been more pronounced had the traditional channel in the experiment been a more significant event—like a West Wing address—rather than a written press release. Moreover, some studies suggest the use of different social media platforms can yield divergent effects during international crises (Narang and Williams 2022). Additional experiments might therefore explore whether threat issuance using different social media platforms affects perceived threat credibility among both decision-makers and the public.

Finally, scholars might explore the generalizability of our findings. Our research design asks respondents in the United States, India, and Singapore to assess the credibility of a threat. Future work might assess whether our results hold among respondents in other states or if different states are named as the issuing and target actors. Pursuing these questions will further enrich our understanding of the interactions between emerging technology and international politics.

References

Barberá, Pablo, and Thomas Zeitzoff. 2018. "The New Public Address System: Why Do World Leaders Adopt Social Media?" *International Studies Quarterly* 62 (1): 121–30. https://doi.org/10.1093/isq/sqx047.

Baum, Matthew A., and Philip B. K. Potter. 2019. "Media, Public Opinion, and Foreign Policy in the Age of Social Media." *The Journal of Politics* 81 (2): 747–56.

Brutger, Ryan, Joshua Kertzer, Jonathan Renshon, Dustin Tingley, and Chagai Weiss. 2021. "Abstraction and Detail in Experimental Design." In .

Carson, Austin, and Keren Yarhi-Milo. 2017. "Covert Communication: The Intelligibility and Credibility of Signaling in Secret." *Security Studies* 26 (1): 124–56. https://doi.org/10.1080/09636412.2017.1243921.

Chesney, Robert, and Danielle Keats Citron. 2018. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." SSRN Scholarly Paper ID 3213954. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.3213954.

Clark, Richard. 2021. "Pool or Duel? Cooperation and Competition Among International Organizations." *International Organization*, 1–21.

Collins, Stephen D., Jeff R. DeWitt, and Rebecca K. LeFebvre. 2019. "Hashtag Diplomacy: Twitter as a Tool for Engaging in Public Diplomacy and Promoting US Foreign Policy." *Place Branding and Public Diplomacy* 15 (2): 78–96. https://doi.org/10.1057/s41254-019-00119-5.

Coppock, Alexander, and Oliver A. McClellan. 2019. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6 (1).

Dafoe, Allan, Baobao Zhang, and Devin Caughey. 2018. "Information Equivalence in Survey Experiments." *Political Analysis* 26 (4): 399–416. https://doi.org/10.1017/pan.2018.9.

Diamond, Jeremy, Caroline Kelly, and Greg Clary. 2020. "Donald Trump Warns Iran That 'we Have Targeted 52 Iranian Sites' If Any Americans Are Attacked." CNN. January 5, 2020. https://www.cnn.com/2020/01/04/politics/trump-warning-iran-52-assets/index.html.

Dietrich, Simone, Heidi Hardt, and Haley J. Swedlund. 2021. "How to Make Elite Experiments Work in International Relations." *European Journal of International Relations*, February. https://doi.org/10.1177/1354066120987891.

Duncombe, Constance. 2017. "Twitter and Transformative Diplomacy: Social Media and Iran–US Relations." *International Affairs* 93 (3): 545–62.

Esper, Mark T. 2022. *A Sacred Oath: Memoirs of a Secretary of Defense During Extraordinary Times*. New York, NY: William Morrow.

Fearon, James. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *The American Political Science Review* 88 (3): 577–92.

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

Fuhrmann, Matthew, and Todd S. Sechser. 2014. "Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence." *American Journal of Political Science* 58 (4): 919–35.

Gelpi, Christopher, and Joseph Grieco. 2015. "Competency Costs in Foreign Affairs: Presidential Performance in International Conflicts and Domestic Legislative Success, 1953–2001." *American Journal of Political Science* 59 (2): 440–56.
Ghitis, Frida. 2019. "This Is What Happens When Trump Makes Foreign Policy by Tweet." Politico Magazine. January 14, 2019. https://politi.co/2FxkZ9L.
Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *The Journal of Conflict Resolution* 46 (2): 175–200.
Hedling, Elsa, and Niklas Bremberg. 2021. "Practice Approaches to the Digital Transformations of Diplomacy: Toward a New Research Agenda." *International Studies Review*, no. viab027 (June). https://doi.org/10.1093/isr/viab027.
Jahng, Mi Rosie, and Jeremy Littau. 2016. "Interacting Is Believing: Interactivity, Social Cue, and Perceptions of Journalistic Credibility on Twitter." *Journalism & Mass Communication Quarterly* 93 (1): 38–58. https://doi.org/10.1177/1077699015606680.
Jervis, Robert. 1970. *The Logic of Images in International Relations*. Princeton, N.J.: Princeton University 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.
Johnson, Thomas J., and Barbara K. Kaye. 2014. "Credibility of Social Network Sites for Political Information Among Politically Interested Internet Users*." *Journal of Computer-Mediated Communication* 19 (4): 957–74. https://doi.org/10.1111/jcc4.12084.
Kertzer, Joshua D. 2016. *Resolve in International Politics*. Princeton, NJ: Princeton University Press.
Kertzer, Joshua D., and Jonathan Renshon. 2022. "Experiments and Surveys on Political Elites." *Annual Review of Political Science* 25.
Kreps, Sarah. 2020. *Social Media and International Relations*. Cambridge: Cambridge University Press.
Kurizaki, Shuhei. 2007. "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy." *The American Political Science Review* 101 (3): 543–58.
Lupton, Danielle L. 2020. *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Ithaca: Cornell University Press.
McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. New York: Signet Books.
McManus, Roseanne W. 2017. *Statements of Resolve: Achieving Coercive Credibility in International Conflict*. New York: Cambridge University Press.
Mir, Asfandyar, Tamar Mitts, and Paul Staniland. 2022. "Political Coalitions and Social Media: Evidence from Pakistan." *Perspectives on Politics*, August, 1–20. https://doi.org/10.1017/S1537592722001931.
Morris, Meredith Ringel, Scott Counts, Asta Roseway, Aaron Hoff, and Julia Schwarz. 2012. "Tweeting Is Believing? Understanding Microblog Credibility Perceptions." In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 441–50. CSCW '12. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2145204.2145274.

Narang, Vipin, and Heather Williams. 2022. “Thermonuclear Twitter.” In *The Fragile Balance of Terror: Deterrence in the New Nuclear Age*, edited by Vipin Narang and Scott D. Sagan, 63–89. Ithaca, N.Y.: Cornell University Press.

Press, Daryl G. 2007. *Calculating Credibility: How Leaders Assess Military Threats*. Cornell Studies in Security Affairs. Ithaca, NY: Cornell University Press.

Schelling, Thomas. 1961. “Experimental Games and Bargaining Theory.” *World Politics* 14 (1): 47–68.

———. 1966. *Arms and Influence*. New Haven, CT: Yale University Press.

Seib, Philip M. 2012. *Real-Time Diplomacy: Politics and Power in the Social Media Era*. 1st ed. New York: Palgrave Macmillan.

Shapiro, Ari, dir. 2018. “Foreign Policy Expert Considers Repercussions Of Trump’s Twitter Diplomacy.” *All Things Considered*. National Public Radio.

Shepp, Jonah. 2018. “How U.S. Foreign Policy Is Being Shaped by Trump’s Tweets.” *Intelligencer*, January 19, 2018. https://nymag.com/intelligencer/2018/01/how-u-s-foreign-policy-is-being-shaped-by-trumps-tweets.html.

Smith, Alastair. 1998. “International Crises and Domestic Politics.” *American Political Science Review* 92 (3): 623–38.

Snyder, Jack, and Erica Borghard. 2011. “The Cost of Empty Threats: A Penny, Not a Pound.” *The American Political Science Review* 105 (3): 437–56.

Statista. 2021. “Twitter Global MDAU 2021.” Statista. 2021. https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/.

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

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

Twiplomacy Study 2020. 2020. Burson Cohn & Wolfe.

Ven Bruusgaard, Kristin, and Jaclyn A. Kerr. 2020. “Crisis Stability and the Impact of the Information Ecosystem.” In *Three Tweets to Midnight*, edited by Harold Trinkunas, Herbert Lin, and Benjamin Loehrke, 137–58. Stanford, CA: Hoover Institution Press.

Westerman, David, Patric R. Spence, and Brandon Van Der Heide. 2012. “A Social Network as Information: The Effect of System Generated Reports of Connectedness on Credibility on Twitter.” *Computers in Human Behavior* 28 (1): 199–206. https://doi.org/10.1016/j.chb.2011.09.001.

Williams, Heather, and Alexi Drew. 2020. *Escalation by Tweet: Managing the New Nuclear Diplomacy*. London: Kings College London.

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. https://doi.org/10.1177/0022002718785693.

Zeitzoff, Thomas. 2017. “How Social Media Is Changing Conflict.” *Journal of Conflict Resolution* 61 (9): 1970–91. https://doi.org/10.1177/0022002717721392.