# Tweets to the Streets? Effects of a Leader's Social Media Messaging on Nationalist Mobilization

Fahd Humayun

Department of Political Science, Tufts University, Medford, USA
Email: fahd.humayun@tufts.edu

**Abstract** Can a leader's use of social media in an external crisis increase domestic nationalist mobilization? In this research note I leverage an original writing simulation on serving national security elites in Pakistan to demonstrate that the vocabulary that leaders employ is endogenous to their messaging platform, with crisis messages composed for Twitter/X communicating greater levels of affective content than similar content composed for official communiqués. I then randomize elite-curated content across an online sample of Pakistani social media users to show that bundling escalatory intent with affective platform-specific vocabulary reduced respondent sensitivity to crisis details, increased public jingoism, and slightly increased domestic willingness to engage in street protest. I discuss the strategic implications of these findings for leaders who, conscious of the mobilizing risks of social media, may wish to avoid tying their hands in a crisis.

In the 218 weeks that former Prime Minister Imran Khan of Pakistan was in power, the cricket star turned politician took to his personal Twitter account on a total of ninety-six occasions to signal his resolve to defend Pakistan against arch-rival India. This averaged roughly one anti-India social media post for every fortnight that Khan was in office.

Neither the content nor tone of the Pakistani prime minister's messages were altogether unsurprising, given the state of hostility that existed between South Asia's two nuclear-armed rival neighbors. In early 2019, India and Pakistan had come precipitously close to war after an attack by militants in the disputed territory of Kashmir killed Indian soldiers, prompting India to undertake an unprecedented airstrike into Pakistan, which it had accused of sponsoring militants.[1] In the months that followed the airstrike, Prime Minister Khan took to social media repeatedly to publicly signal Pakistan's resolve to defend itself from further cross-border aggression. These posts swiftly amassed thousands of likes and reshares.

Imran Khan is not the only leader in recent years to have employed their personal social media account to signal resolve against an external enemy in or around an interstate crisis. During border clashes between Thailand and Cambodia in 2025,

1. "Balakot: Indian Air Strikes Target Militants in Pakistan," BBC News, 26 February 2019.

*International Organization* 80, Spring 2026, pp. 407–430
© The Author(s), 2026. Published by Cambridge University Press on behalf of The IO Foundation.
doi:10.1017/S0020818326101362

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

Cambodia's President Hun Sen actively threatened Thailand and rallied the Cambodian public using Facebook. In Iran, the late Ayatollah routinely took to social media to warn both Israel and the United States of the consequences of an attack. And Azerbaijan's president has more than once threatened war with Armenia via lengthy tweets. Given their significant reach and capacity for instant visibility, it is not difficult to see why leaders' social media accounts have become an indispensable new-age tool of interstate conflict, allowing leaders to advertise their resolve publicly and online, in a bid to, at least theoretically, deter rival countries from unwanted aggression.[2] The logic on which this form of deterrence rests, moreover, intimately pivots on a leader's domestic audience and the assumption that if a leader makes empty threats, they may compromise their popularity at home.[3]

But despite the growing propensity of world leaders to use their personal social media accounts to communicate resolve, the domestic consequences of their chosen communication medium remain poorly understood. Is a leader's social media messaging just as effective in generating domestic nationalist mobilization as more conventional forms of government communication, such as official statements or government-issued communiqués? Can a leader's social media shape domestic threat perceptions in ways that are similar to how more official communication might? What, if any, are other consequences of a leader's choice of communication medium on escalation dynamics?

This research note attempts to advance our understanding of the impact of leaders' social media on domestic nationalist mobilization in the highly relevant context of the India–Pakistan rivalry. Nuclear-armed neighbors, India and Pakistan have fought four hot wars and engaged in several precipitous near-wars, the most recent of which in May 2025 was contained only on account of last-minute US intervention. With the risk of nuclear escalation carrying consequences for over 1.3 billion people, roughly a fifth of the world's population, South Asia's stability remains a critical US national security interest.

This note shows that *how* a leader chooses to phrase their message to an external rival in an interstate crisis is, in fact, endogenous to their choice of messaging platform. Specifically, I show that crisis messages conveyed via a leader's social media account are more likely to employ political affect via tone and sentiment; this in turn has the potential to heighten domestic out-group animosity and generate modestly stronger public rallies than more conventional forms of government communication. Consequently, because social-media-induced domestic mobilization signals credibility and resolve vis-à-vis an external adversary, I suggest that leaders seriously contemplating military force may, in fact, need to be cautious about signaling escalatory intent via social media, lest it unnecessarily tie their hands around the use of force.

2. Harris and Lin-Greenberg 2024; Jablanski, Lin, and Trinkunas 2024; Narang and Williams 2023; Schultz 2017.
3. Fearon 1994, 1997; Schultz 1998, 2001; Weeks 2008; Weiss 2014.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

I sequence these arguments in three steps. First, to demonstrate the selection effects around language when it comes to social media message composition, I pair observational text analysis of real-world crisis messages with a novel writing simulation involving serving Pakistani national security elites to demonstrate that elite messaging during foreign policy crises communicated higher levels of affective content when it was drafted for a leader's personal social media account than when it was drafted for official communiqués. Next, I use these real-world, elite-sourced messages as the basis for four vignettes in a survey experiment fielded on an online panel of 1,687 Pakistani social media users. I show that bundling escalatory intent with the less affective language of official communiqués heightened respondents' attentiveness to crisis details, while bundling escalatory intent with the more affective, platform-specific vocabulary of social media reduced social media users' attentiveness to crisis details as well as nonpartisan objectivity, despite the shorter and ostensibly more digestible form of content presented. I suggest that it was this reduced sensitivity to crisis details, borne of a leader's employment of affect, that blurred the risks of crisis escalation, ultimately resulting in the slight increase in respondent willingness to participate in nationalist anti-foreign street protests. Finally, I discuss the strategic implications of these findings for leaders beyond just South Asia. These arguments, I offer, contribute to a growing and important line of research on the role of communication technologies on both crisis signaling and democratic accountability around the use of force.

## Leaders, Social Media, and Interstate Crises

Nearly every world leader today has an active social media presence, either through personal accounts, official institutional handles, or often both. Leaders are also increasingly turning to these platforms in interstate crises, when they may have inflated incentives to publicly issue threats or make declarations of resolve against external adversaries. In democratic polities, the reason these declarations of resolve are considered credible by adversaries is that they are made in the public domain, since domestic audiences may impose penalties on a signaler who makes empty threats, or backs down.[4] The role of open informational environments is central to this logic of accountability-borne credibility: it has previously been argued, for instance, that audience costs theory implicitly required a free press, because without it the public had no reliable means of obtaining information about the success or failure of a leader's foreign policy.[5] Today, social media is considered a critical component of open information environments. Yet despite an established literature on the domestic drivers of nationalist anti-foreign mobilization we still do not fully know whether a leader's choice of communication medium affects the likelihood and intensity of domestic mobilization.[6]

4. Baum 2004; Croco 2011; Fearon 1994; Guisinger and Smith 2002; Schelling 1966; Tomz 2007.
5. Baum and Groeling 2010.
6. Butt 2016; Debs and Weiss 2016; Weiss 2014; Weiss and Dafoe 2019.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

It is, of course, possible to speculate. On the one hand, we might expect that because social media offers leaders a direct channel of communication with domestic audiences writ large, crisis communication via social media ought to generate considerably stronger internal "rally-'round-the-flag" effects, shoring up support for the leader and mobilizing domestic audiences around the government.[7] At the same time, research on the effects of political communication on partisanship suggests that siloed information distribution patterns unique to social media platforms limit exposure to alternative viewpoints, making it easier for consumers to self-select into streams that more closely match their personal preferences, and likelier for misinformation to enter the system. One potential effect thereof is that social media consumption might reduce the likelihood that an incumbent's supporters will hold their leader to account for poor policy decisions or objectively update their preferences in the face of crisis heuristics such as casualties or elite dissent.[8]

### *Affective Lexicons, Selection Effects*

While the fragmented and siloed nature of social media, with its competing information streams and reduced user exposure to alternative viewpoints, offers one explanation for reduced consumer objectivity, there are other aspects of social media that might potentially make it particularly conducive to mobilizing not just "digital tribes" but actual street protests, especially during an interstate crisis.[9] There is reason to believe, for instance, that the informal and unfiltered nature of social media in fact encourages the formulation of more *affective*, valenced messages than official government-issued statements. A growing body of research on the role of *affect*—sentiment, emotions, feelings, and their role in shaping political behavior—has gained significant traction in international relations (IR) in the past two decades.[10] These works have broadly attempted to conceptualize affect as "general valence feelings toward something," or more specifically as "the positive and negative feelings evoked by a stimulus."[11] At the same time, research on political communication has shown that social networking platforms in fact encourage leaders to adopt not only more aggressive and inflammatory rhetoric, but also to rhetorically exaggerate war claims or escalatory moves, which, "in theory, [ought to] take the form of sober, tempered, no-nonsense communications designed to deter or compel an adversary … in practice, however, this flavor of rhetorical escalation constitutes the exception rather than the rule."[12] Research has also found that social

7. During a crisis involving a foreign actor, domestic dissent is unlikely due to social psychological dynamics that lead to increased domestic cohesion, which is why international threats generate a surge of patriotism and increased popular support for the executive. See Coser 1956; Mueller 1973; Russett 1993.
8. Baum and Groeling 2008; Baum and Potter 2019; Levy 2021; Stroud 2010.
9. O'Loughlin 2024.
10. For example, Hymans 2006 was among the first contemporary works to explicitly link two specific emotions felt by decision makers, fear and pride, to a key foreign policy outcome: nuclear proliferation. See also Eznack 2011, 2013; Hutchison and Bleiker 2014; Mercer 2014; Sasley 2010.
11. Betsch 2005, 41.
12. Greenhill 2024, 118.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

media algorithms routinely amplify charged content, and messages that express anger, out-group animosity, and "us versus them" narratives, all tailored for fast, emotional processing, are more likely to be communicated via social media in the first place.[13] These selection effects are not inconsequential; it has been argued, for instance, that a leader's decision to communicate with their public via social media platforms is not only deeply strategic, but that "the choice of language that elites use can influence other behavior … motivate followers, set an agenda, and build a sustainable coalition."[14]

As an initial probe into the extent to which real-world elite rhetoric indeed differs consequentially across online and offline settings, I analyzed 125 Twitter threads and press releases issued by Pakistani Prime Minister Imran Khan during his three-and-a-half years in office, and that specifically centered on arch-rival India. The resulting corpus of documents (≈17,000 words) was analyzed using Linguistic Inquiry and Word Count (LIWC) sentiment-analysis, a recently updated text analysis tool that calculates the proportion of words in psychologically meaningful categories, including affective language, tone or sentiment (positive/negative), and emotional valence (positive/negative).[15] Recent IR research, for instance, has applied LIWC to study the public's emotional response to terrorism, variation in sentiment across UN General Assembly statements, and the deployment of "moral" language in social media and policy discourse during the Russia–Ukraine conflict.[16]

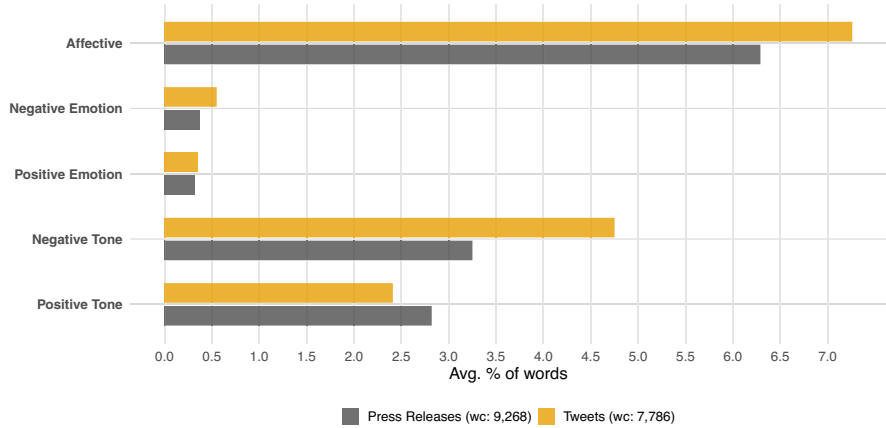
Pertinently for this investigation, LIWC-22's dictionary treats "affect" as a broad psychosocial construct that includes but also distinguishes between (1) broad *tone* or *sentiment* (conveyed by words like "good," "well," "bad," "wrong," "too much," as well as contextually charged words such as "beautiful," "kill," "funeral") and (2) targeted *emotion categories* with explicit emotion labels (such as "happy," "hate," "hurt," "tired," "frustrated.")[17] This distinction is significant because it means that the positive or negative feelings invoked by affective stimuli need not be driven solely by explicit emotion words, and that in practice, affect can also be conveyed through sentiment and tone that might still signal emotional engagement, albeit in the absence of clearly valenced emotional labels.

Analysis of LIWC-derived linguistic differences across Prime Minister Khan's tweets and press releases as captured by Figure 1 shows that on average, 7.36 percent of the total vocabulary used for tweets was classified as affective content, compared to 6.2 percent of the total vocabulary used for press releases. This difference, furthermore, was driven primarily by shifts in tone and sentiment, rather than by increases in explicit positive or negative emotion words. Relative to press releases, tweets were less likely to contain longer words, more likely to use language with a negative tone (4.75 percent compared to 3.25 percent), less likely to use language conveying positive tone (2.4 percent compared to 2.8 percent), and a little less likely

13. Iyengar, Sood, and Lelkes 2012; McDermott 2024; Milli et al. 2023.
14. Barberá et al. 2022, 50.
15. Boyd 2017; Pennebaker 2011.
16. Amjadi and John 2025; Mesquita 2024; Wayne 2023.
17. Boyd et al. 2022.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

to exhibit markers of analytical thinking—that is, language associated with formal, logical, and hierarchical patterns of expression.[18]




*Note*: On average, tweets (yellow) contained greater levels of affect than press releases (grey).

**FIGURE 1.** *Affective content across tweets and press releases according to LIWC's language categories*

Based on these differences and building on claims in the literature surveyed earlier, I theorize that when using social media, leaders ought to use more affective language capable of evoking positive or negative feelings among audiences. From this I offer:

*H1: In an interstate crisis, a leader's social media messages should be more likely to communicate affective content than messages conveyed via more traditional channels, such as a government-issued press release.*

## Crowdsourcing Social Media Content From Real-World National Security Elites

To test this initial hypothesis, in the summer of 2020 I recruited real-world serving national security practitioners from Pakistan in a short writing simulation with the specific objective of curating a directory of elite-sourced content that could later be used as the basis for primes for a larger survey experiment.[19] As part of the writing simulation, the group of national security practitioners was provided a hypothetical

18. For complete scores as well as notes on data sourcing, see section A.1 of the appendix.

19. An illustration of this two-stage research design, showing exactly how national security elite-sourced content fed into the subsequent online survey experiment, is provided in section B.2 of the appendix. For more details on recruitment and participation, see section B.1 of the appendix.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

foreign policy crisis prompt involving Pakistan's arch-rival India, and asked to draft content for a short press statement that could be issued by the prime minister's office for immediate release to news channels and the morning papers, as well as a short social media thread for the prime minister to tweet from his personal social media account.[6]

Pakistan presents a useful test case to study variation in the medium of crisis communication—tweets versus official communiqués—on emotional activation and nationalist mobilization. Not only is it an example of a non-Western country where social media usage is rapidly expanding, but it is also a nuclear-armed state with a history of repeated crises with India, many of which have escalated rapidly against a backdrop of explicit and implicit elite signaling.[20] The India–Pakistan rivalry too remains a relatively understudied case in IR, despite its proclivity for conflict with direct implications for a fifth of the world's population.[21]

The selection criteria for the national security practitioners who participated in the writing simulation was twofold: in addition to serving in official capacities, practitioners were chosen if they had previously managed official or quasi-official government social media handles, and been involved in the drafting of real-world institutional press releases. The recruited experts were provided text-entry boxes corresponding to the real-world character limit that was (at the time) a feature of Twitter/X's interface; however, they could add more tweets if they wished to compose a Twitter/X "thread" (a series of connected tweets). There was no character constraint on the text-entry boxes for press statements. Participants were asked to pay close attention to the *language* and *tone* they would expect of a press statement issued by the prime minister's office, compared to what they would anticipate from the prime minister's Twitter feed in the exact same crisis. The twenty-six messages that were produced are reproduced in full in section B.4 of the appendix.

Three research assistants were subsequently and independently tasked with employing axial coding to identify common themes and attributes across the two categories of content in a bid to parse out key differences. Axial coding is an established qualitative data analysis technique that combines inductive and deductive reasoning to derive a set of broader thematic categories from given segments of text.[22] I provide a detailed account of how axial coding was employed to identify recurring themes from the expert-authored content in section B.5 of the appendix. Table 1 presents a sample tweet and press release authored by the same national security author in the writing exercise, along with platform-linked framing differences based on the codes and themes derived across the twenty-six original messages produced as part of the exercise.

20. Butt 2016; Cohen 2002; Haqqani 2005; Mir and Siddiqui 2024.

21. Bisaria 2024; Clary, Lalwani, and Siddiqui 2021; Cohen 2002; Ganguly 2002.

22. Descriptive codes summarize explicit content, while interpretive codes delve into underlying meanings and connections. For example, in a corpus of written content simulating national security crisis deliberations, a descriptive code might be "loss of life," while an interpretive code could be "national tragedy." See also Glaser and Strauss 1967; Simmons 2017.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press



**TABLE 1.** *Sample tweet (top) and press release (bottom) authored by the same national security practitioner, illustrating platform-linked framing differences*

| | Descriptive attributes | | | Substantive attributes | |
|---|---|---|---|---|---|
| *Authored content sample* | *Form* | *Personalization* | *Losses* | *Retaliation* | *Rally-'round-the-flag* |
| Tweet:<br>India committed an unprovoked act of aggression against Pakistan last night. I assure my countrymen; Pakistan will respond at a time and place of our choosing. I call on the country to unite behind our armed forces. Insha'Allah Pakistan will prevail. | Compressed; conversational; affective/ abbreviated wording; occasional errors. | Leader-centric ("I," "my government;") concentrated agency. | Affective/emotive treatment of casualties (e.g., "*jawans*," and to a lesser extent, "martyrs.")[23] | Emphasis on resolve, honor, risk, and urgency; opponent-centered framing. | Direct, emotive appeals for unity; occasional anthropomorphizing of the homeland. |
| Press release:<br>The National Security Committee convened under the Chairmanship of the Prime Minister on Sunday. The meeting was attended by federal ministers, national security adviser, all service chiefs, and intelligence heads. The Prime Minister was provided a detailed briefing on the unprovoked attack by India against Pakistan resulting in 13 casualties, including four civilians martyred. Pakistan considers the attack a blatant violation of international law and Pakistan's sovereignty … Pakistan retains its right to respond to India's unprovoked attack at a time and place of its choosing. | Formal, official register; complete sentences. | Institutional voice ("it was decided;") diffuse agency (emphasis on military as consultative stakeholder). | More prosaic numerical reporting; minimal emotive framing. | Emphasis on fairness, norms, legality and collective responsibility; assesment/ procedure stressed. | Appeals for unity framed institutionally; fewer direct public appeals. |

*Note:* The columns record broader platform-linked differences in tone, personalization, and attribution, identified using axial coding across the twenty-six original messages sourced in the simulation.

23. *Jawan* is a South Asian term (both in Hindi and Urdu) that translates roughly as "soldier" or "rank-and-file serviceman," but carries a colloquial and relational connotation, akin to "our soldiers" or "the boys" than to a more neutral bureaucratic label like "troops "or "personnel." To this end, it can be considered more contextually charged than "martyr," which may be slightly less intensely charged, but still charged nonetheless.

The three human coders broadly uncovered two descriptive dimensions (related to form and tone, and personalization), and three substantive dimensions (related to loss of life, retaliation, and mobilization) across the two categories of texts. Platform-linked differences across these dimensions together suggest that elite-authored content for social media was more likely to employ political affect capable of evoking positive or negative feelings among readers. For instance, in describing casualties, content authored for social media was more likely to refer to the loss of "*jawans*" (young men), and more likely to frame the necessity of retaliation around ideas of national honor and resolve. Press releases were relatively less contextually charged when discussing casualty counts, and more likely to invoke international legality and due process when discussing potential retaliation. Furthermore, it was not inconceivable that these attributes also interacted to co-produce greater affective intensity: for instance, the shorter length and informality of tweets made affective language appear more concentrated, while the personal emphasis on the prime minister both simplified attribution and, at least in appearance, reduced situational complexity.

Based on these results, I theorized that affective messages delivered to domestic audiences could potentially result in several consequential outcomes, ranging from heightened domestic out-group animosity, to increased political tribalism and reduced receptivity to disconfirming or counter-attitudinal information.[24] Furthermore, the unique ways in which urgency and the visibility of deliberation and military consultation was communicated across social media messages could plausibly serve to reduce sensitivity to neutral, evaluative cues, as well as respondent attentiveness to the risks of crisis escalation.[25] For these reasons, and building on existing scholarship on social media and partisanship, I theorized that *because* a leader's social media messaging in an interstate crisis was likely to bundle escalatory intent with political affect, this messaging should be more likely to generate domestic public mobilization. From these arguments, I derived the following additional hypotheses:

*H2: The effect of escalatory messaging on reducing respondent attentiveness will be greater when the message is delivered via social media than via a traditional government-issued press release.*

*H3: The effect of escalatory messaging on reducing respondent receptivity to nonpartisan elite cues will be greater when the message is delivered via social media than via a traditional government-issued press release.*

*H4: The effect of escalatory messaging on hardening respondent attitudes toward an external adversary will be greater when the message is delivered via social media than via a traditional government-issued press release.*

24. Baum and Potter 2019; Kreps 2020; O'Loughlin 2024; Schultz 2017; Stroud 2010.

25. Kenwick and Maxey 2022 have shown that civilian leaders often have strategic incentives to choreograph how they publicly reference "the military's advice" to increase support for risky action as well as support for the leader in question.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

*H5: The effect of escalatory messaging on respondents' appetite for retaliation, jingoism, and willingness to protest will be greater when the message is delivered via social media than via a traditional government-issued press release.*

## Experimental Evidence from Pakistani Social Media Users

To formally test these relationships, and using experimental vignettes whose language was informed by the crisis content produced in the writing exercise, I proceeded to field a survey experiment in the summer of 2022 on a representative sample of 1,687 Pakistan-based adult social media users.[26]

Respondents were given the option to take the survey in either Urdu or in English.[27] Three linear models were preregistered to estimate the effects of escalatory intent and communication medium on the specified outcomes: $Y_i = \beta_0 + \beta_1$ Escalatory (model 1), where the parameter of interest $\beta_1$ was the marginal effect of reading an escalatory message; $Y_i = \beta_0 + \beta_2$ tweet (model 2) where the parameter of interest $\beta_2$ was the marginal effect of the medium of communication (specifically, whether the statement was delivered via Twitter); and $Y_i = \beta_0 + \beta_1$ Escalatory $+ \beta_2$ tweet $+ \beta_3$ (Escalatory $\times$ Tweet) (model 3) where the parameter of interest $\beta_3$ was an interaction term that captured the *conditional* average treatment effect (CATE), that is, whether the effect of an escalatory message differed by communication medium.

The survey began by asking respondents to select their political affiliation, before introducing a hypothetical foreign policy crisis centered on an Indian cross-border attack that had resulted in the deaths of five Pakistani soldiers and five Pakistani civilians, including women and children. The exact vignette read by all respondents was as follows:

> Please read the following scenario carefully. India announces that it has carried out surgical strikes against Pakistan. The Government of Pakistan confirms that India has undertaken missile attacks that have martyred 5 Pakistani soldiers and 5 Pakistani civilians including women and children. Pakistan terms India's attack a violation of its sovereignty and an act of war. On the next screen you will see a [TWEET/PRESS RELEASE] issued by the Prime Minister.

In a 2x2 design, the survey then randomized whether a Pakistani prime minister from a respondent's preferred political party responded to the provocation with an *escalatory* or *de-escalatory* message, and whether they communicated this message

26. The survey was fielded two years before Twitter was temporarily blocked in Pakistan from February 2024 to May 2025.

27. The survey was administered on Qualtrics and fielded on respondents registered with Cint, an international consumer network for digital market research. Respondents were deemed eligible to participate in the survey if they were based in Pakistan and self-reported as routinely using social media. Eligibility was further restricted to respondents who were able to pass a basic attention check at the start of the survey. I include the original survey instruments in both languages in sections C.2 and C.3 of the appendix.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

using *Twitter*, or by issuing *a press release*.[28] Crucially, each of the four experimental vignettes to which respondents could be randomly assigned was designed keeping in mind the observational characteristics of the elite-authored messages summarized in Table 1 earlier. For example, the Twitter treatments incorporated more affectively charged references to casualties (for example, "martyred," "*jawans*"), greater rhetorical emphases on retaliation, resolve, and urgency, more opponent-centered framing, and direct rallying appeals that anthropomorphized the nation. The shorter length and informal register of tweets also made this affective content appear more concentrated, while the use of personalized phrasing (for example, "I assure my countrymen") was intentionally employed to make attribution to the leader more direct, while reducing situational complexity. By contrast, the press release treatments reflected the more formal, institutionally anchored language typical of official communiqués. While the full treatments are reproduced in Figures A.6 and A.7 in section B.6 of the appendix, Figure 2 below broadly summarizes the survey experiment's design.

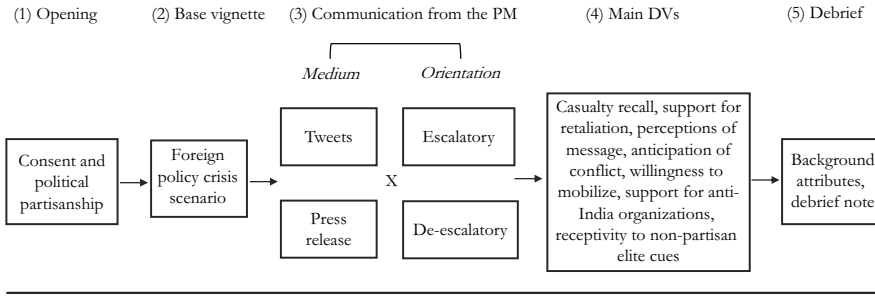



**FIGURE 2.** *Survey design*

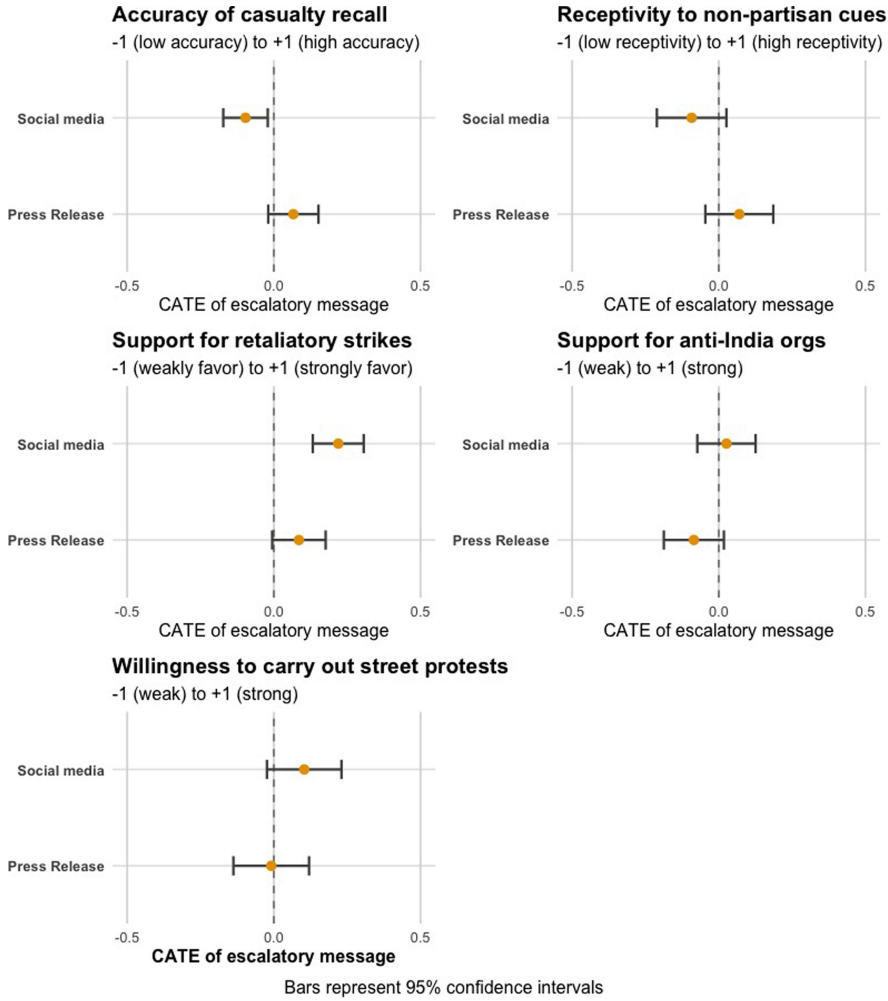
### *Experimental Results*

Analysis of the experiment's main conditional treatment effects suggests that the effect of escalatory messaging on respondents' appetite for retaliation, jingoism, and willingness to protest was greater when the message was delivered via a leader's social media than via a traditional government-issued press release, while escalatory messaging delivered via social media just slightly dampened the extent to which respondents thought escalation to war was likely, possibly suggesting that it may have obscured the risks of actual crisis escalation. Figure 3 shows the conditional average treatment effects (CATEs) of an escalatory message when delivered via a leader's tweets versus a leader's press release on preregistered outcomes of interest.[29]

28. This design closely approximates that used by Harris and Lin-Greenberg 2024 who manipulate whether a US president issued a threat to a rival using Twitter or a White House press release.

29. I include summary statistics and difference-in-means estimates in section C.5 of the appendix.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press




*Notes*: Estimated separately for when the medium was social media, versus a press release. Points represent the estimated effect of receiving an escalatory message within each medium. See the appendix for numerical estimates.

**FIGURE 3.** *Conditional average treatment effects of escalatory messaging*

### *Respondent Attentiveness*

The top left panel in Figure 3 reports respondents' attention to crisis details (specifically, their recall of the number of casualties they had read in the scenario) as a consequence of the medium by which the prime minister's message was conveyed. Responses were recoded −1 if respondents inaccurately recalled the number of

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

casualties in the scenario they had read, and +1 for accurate recall.[30] The effect of reading an escalatory message on respondents' accuracy of recall when the prime minister issued a press release was 0.067 (se = 0.044), but the effect of issuing an escalatory message when the prime minister took to Twitter was −0.096 (se = 0.038).[31] With respect to H2, these results suggest some reduced attention to detail when respondents were exposed to escalatory tweets, despite the shorter and more digestible form of complex information, compared to escalatory press releases.

*Receptivity to Nonpartisan Elite Cues*

The top right plot, meanwhile, captures respondents' receptivity to nonpartisan elite cues, which I use as a measure of respondent sensitivity to neutral, evaluative information that might also be counter-attitudinal during a high-stakes crisis. In the actual survey, respondents were told that some members of the national security council (a nonpartisan executive body) had expressed reservations regarding the prime minister's statement. Respondents were asked whether these elite, nonpartisan reservations ought to be taken into consideration, on a scale of −1 (did not need to be taken into consideration) to +1 (needed to be taken into consideration). Results show that the effect of an escalatory message on respondents' receptivity to nonpartisan cues when the leader's medium of communication was a press release was 0.07 (se = 0.059), while the effect of an escalatory message when the medium was Twitter was −0.093 (se = 0.06) (see Table A.10 in the appendix for difference-in-means estimates). In keeping with the literature on the effect of social media on digital tribalism, these results do seem to tentatively suggest a possible reduction in receptivity to nonpartisan elite cues when a leader expressed escalatory intent via their social media, in contrast to a very slight increase in receptivity to nonpartisan elite cues when the leader's intent was conveyed in a press release.

*Support for Escalation*

Next, to what extent did treatment shape respondents' support for crisis escalation? I measure support for crisis escalation both directly, by asking respondents if they supported retaliatory military strikes, and indirectly, by asking respondents how willing they would be to endorse anti-India religious-militant outfits. Responses to both questions were again coded on a scale from −1 to +1, with higher values corresponding to greater respondent hostility. The center left panel in Figure 3 shows that the effect of an escalatory message on respondent support for retaliatory strikes

30. For ease of interpretation and comparability across items, all outcome variables were, per preregistration, rescaled from −1 to +1. For example, a three-category variable originally coded 1, 2, 3 became −1, 0, +1. A binary variable coded 0, 1 (such as for accuracy of casualty recall) became −1, +1. As a robustness check, main results were replicated using the original (unrescaled) codings, with substantively identical results (see appendix).

31. Note, these estimates are derived from model (3) in Table A.9 in the appendix, where the CATE of issuing an escalatory message when the medium was social media is calculated as the sum of the coefficients on Escalatory ($\beta_1$) and the interaction term Escalatory × Tweet ($\beta_3$).

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

against the enemy country, when communicated by Twitter, was 0.22 (se = 0.04, $p < .001$), while the effect of an escalatory message communicated by press release was only 0.08 (se = 0.04) (see Table A.11 in the appendix for difference-in-means estimates). These results modestly favor H4. Meanwhile, the center right panel shows that, when asked if they believed that the country should support anti-India organizations, the effect of seeing a leader's tweets conditional on an escalatory statement was negligible, but the effect of escalation when the medium was a press release was −0.08 (se = 0.05, $p = 0.1$), suggesting that an escalatory statement via press release very slightly reduced out-group animosity (as measured by respondents' endorsement of anti-India organizations). Taken together, these results offer some potential evidence in support of H5, which held that a leader's escalatory messages, when delivered via Twitter, should enhance support for retaliation and/or public jingoism.

*Willingness to Domestically Mobilize*

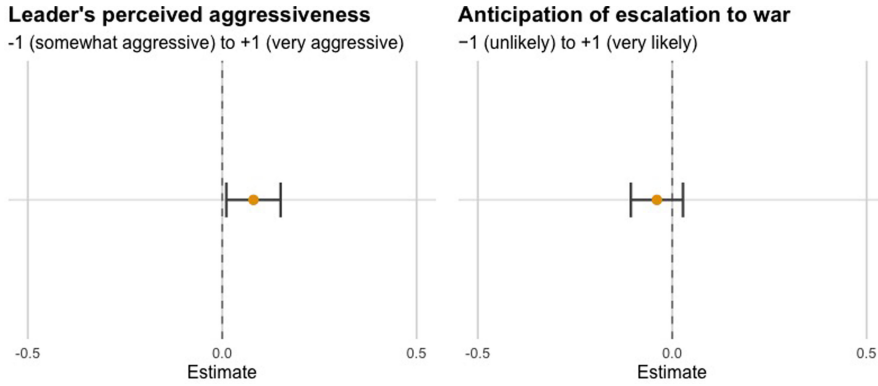
The bottom left panel in Figure 3 indicates respondents' reported willingness to participate in nationalist, anti-foreign mobilization, and shows that escalation had little to no effect on citizen mobilization when the medium of communication was a press release, but had a modest effect of 0.104 (se = 0.06) when the medium was Twitter (see Appendix Table A.13 for numerical tables). When considered in tandem with the earlier findings, these results offer some tentative evidence that a leader's escalatory intent expressed on social media was somewhat more successful in both eliciting public jingoism and modestly more successful in mobilizing citizens in nationalist, anti-foreign protests, than when expressed via press release. The pathways for these attitudinal shifts, the results also suggest, may well have been affect-driven tribalism, accompanied by increased out-group animosity, reduced receptivity to nonpartisan cues, and reduced attentiveness to detail among respondents who read a leader's escalatory messages via social media.

*Perception of Likelihood of Conflict*

In addition to reducing respondents' attentiveness to detail and increasing out-group animosity and support for escalation, did bundling escalatory messaging with the platform-specific vocabulary of social media also obscure the risks associated with crisis escalation? Two additional results to this end are worth highlighting in Figure 4. First, Figure 4's left-hand panel shows that the *marginal* effect of reading a leader's tweets on the *perceived aggressiveness* of the leader's escalatory message was 0.08 (se = 0.035). Meanwhile, the right-hand panel shows the marginal effect of reading a leader's tweets on respondents' *perceived likelihood of war* was −0.039 (se = 0.034). These results, I offer, provide some additional evidence that a leader's social media posts, despite their greater perceived levels of aggressiveness, very slightly dampened the extent to which respondents thought escalation to war was likely, possibly indicating that escalatory messaging on Twitter may have obscured the risks of crisis

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

escalation. If correct, this would align with respondents' comfort in engaging in public jingoism as well as their greater readiness to mobilize when a leader's messages were communicated via social media (see Tables A.14 and A.15 in the appendix for numerical tables).




*Note*: See the appendix for numerical tables and difference-in-means estimates.

**FIGURE 4.** *Marginal effects of a leader's tweets on respondents' perceptions of the aggressiveness of the message (L) and anticipation that the crisis might escalate to war (R)*

These preliminary conclusions, suggesting that escalatory intent conveyed via a leader's social media may have indeed heightened out-group hostility, jingoistic sentiment, and willingness to mobilize among survey takers, are not at odds with findings in neural science that demonstrate that the brain uses emotion as part of the decision-making process, and that exposure to affective language has the ability to influence an audience's subsequent political choices. They also offer an alternative account to recent work on crisis signaling and social media by Harris and Lin-Greenberg, which finds no significant difference in the externally perceived credibility of threats issued via a leader's tweets versus formally issued press releases.[32] While Harris and Lin-Greenberg focused on the interpretation of a leader's threats by foreign observers, it nonetheless rested on the assumption that foreign perceptions of a leader's resolve are informed by a mobilized domestic public square. The findings in this research note seek to clarify the mechanisms by which a leader's social media messages influence domestic audiences in a crisis in the first place, showing that a leader's escalatory online rhetoric may potentially obscure some of the risks of crisis escalation, while paradoxically inducing affect-driven nationalist mobilization.

32. Harris and Lin-Greenberg 2024.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

Finally, two important limitations of the survey experiment, and the inferences that can be drawn from it, warrant mention. The first pertains to treatment design. While it is indeed the case that the experimental treatments in this section marked a departure from Harris and Lin-Greenberg by bundling both the medium of communication and the language typically used within that medium, this choice was not unintentional: in an effort to mirror real-world behavior to the greatest extent possible, the treatments were designed with a view to ensuring realism and ecological validity by maintaining the real-world rhetorical styles associated with different platforms. However, this design choice comes with two methodological trade-offs. Because medium and linguistic style vary simultaneously, any observed differences in treatment effects could plausibly stem from either or both, limiting the ability to isolate the causal effect of medium alone, even if it more faithfully reflects the environment in which audiences typically receive elite cues. A related confounder could be that a leader's intended principal audience might also well be platform specific, with social media statements aimed at mass publics, and communiques geared toward domestic or international elites. If this is the case, then medium might also be correlated with expected audience, and not just with rhetorical form.

Second, with respect to the question of generalizing from social media users to a broader population of potential nonusers, it is important to acknowledge that the demographic profile of the survey respondents (reported in the appendix) differed systematically from that of the national population.[33] Twitter/X usage in Pakistan, while growing, still remains limited; some estimates suggest that fewer than 5 percent of Pakistanis maintain Twitter accounts, though a recent nationally representative phone survey suggested that approximately 17 percent of respondents reported using the platform.[34] Therefore the sample in this research note, arguably, captures only a politically attentive, digitally connected subset of the broader population.

That said, there are at least three compelling reasons that the dynamics observed here may nonetheless still scale to broader mass public. One, in open and highly porous information environments (such as Pakistan's), exposure to elite social media messaging is not confined to official platform users. Because traditional broadcast and print outlets routinely report on and rebroadcast elite social media statements, online messages get high levels of offline traction.[35] Two, social media messages from the country's leaders in the past have had had demonstrably sizable, real-world mobilizational effects on the broader public. In 2019, Prime Minister Imran Khan used social media to call for a nationwide "Kashmir Hour," which generated coordinated street demonstrations and visible public signaling. Such episodes illustrate how online elite cues can cascade into broader mass participation by citizens

33. Over 70 percent of respondents who took the survey, for instance, reported holding a graduate or postgraduate degree. See Table A.6 for sample characteristics in the appendix.

34. Mir and Siddiqui 2024.

35. Research suggests that while roughly half of Pakistan continues to rely primarily on television and traditional media for news, Internet and social media sources are rapidly catching up, serving as principal news sources for growing segments of the population. See Ramsha Jahangir, "Riding the Social Media Tiger," Dawn News, 5 June 2022. Available at <https://www.dawn.com/news/1693238>.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

who might not themselves directly use social media. Finally, the potential reach of social media across South Asia is rapidly expanding. In 2023, Pakistan had 124 million broadband users and 187 million mobile subscribers, reflecting a sharp growth in the country's digital penetration. This expansion is especially pronounced among younger Pakistanis, who constitute over 60 percent of the population, and whose political awareness and mobilization are increasingly shaped by digital platforms, including Twitter/X.[36] As these cohorts age into greater political relevance, the influence of elite online messaging may become even more consequential for crisis politics.

## Mobilization, Credibility, and Crisis Behavior: South Asia and Beyond

"What if the Cuban Missile Crisis had taken place in today's global information environment, characterized by the emergence of social media as a major force amplifying the effects of information on both leaders and citizens?"[37] While speculating on counterfactuals is never easy, the India–Pakistan rivalry, I offer, presents a contemporary case of a crisis-prone adversarial dyad in both the nuclear and digital age, and thus a useful backdrop against which to test the impact of a leader's social media crisis signaling on support for domestic rallies at home.

The experimental results contained in this research note suggest that escalatory messaging delivered via a leader's social media generated tentatively stronger support among respondents for domestic nationalist mobilization, compared to the more restrained and prosaic language of a standard press release from the leader's office. I argued that this effect stemmed from the use of affective vocabulary germane to social media, which reduced respondent objectivity, dampened attentiveness, and obscured the risks of crisis escalation for domestic audiences.

The implications of these findings, I hazard, create somewhat of a dilemma for crisis stability. If domestic audiences are prone to rally around a leader's social media messaging, and if such mobilization helps signal credibility and resolve to external adversaries, might leaders who are actually contemplating escalation refrain from signaling intent via personal social media accounts, for fear of unnecessarily tying their hands and limiting their ability to pursue de-escalatory off-ramps once escalation is underway?

A brief chronology of the intense crisis-signaling undertaken by both India and Pakistan during the February 2019 Pulwama–Balakot crisis, which very nearly escalated to full war between the night of 25 February and morning of 27 February 2019, offers some corroborative evidence that the two nuclear-armed countries' leaders, faced with very real dangers of crisis escalation, and aware of the inflammatory potential of social media messaging, made deliberate and conscious

36. Hayat, Zahra, and Ali 2022.
37. Jablanski, Lin, and Trinkunas 2024, 1.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

decisions to not use social media to escalate tensions.[38] Instead, crisis signaling by both sides was employed through more conventional channels and platforms, potentially indicating the seriousness with which the issue of escalation was being deliberated, as well as the likely responses in case of any specific military action by one side against the other. Prime Minister Modi's final tweet on the Pulwama attack during the crisis window was on 19 February, a full week before India's retaliatory strike against Pakistan. In that tweet he was vague and nondescript at best in terms of explicit signaling, saying only, "Those who have perpetrated this dastardly act will pay a heavy price" — but did not name Pakistan.[39] Prime Minister Khan's tweets, in the days following this veiled threat on 19 February and India's retaliatory strike on 26 February, continuously emphasized the need for dialogue and de-escalation, rather than trying to incite domestic mobilization.

The Pulwama-Balakot case also sheds some light on the possible real-world data-generating processes behind a leader's choice of communication medium. While data collection for the observational analysis carried out at the outset of the paper showed that official communiqués were often accompanied by parallel messaging via a leader's social media, the 2019 crisis showed that this pattern was *not* universal. During key moments of the crisis, the Pakistani prime minister's office issued formal communiqués while the prime minister's personal Twitter account remained silent, suggesting that leaders do indeed sometimes choose institutional channels without parallel personal social media messaging, and that such choices appear to track political context and strategic incentives—for instance, during moments in which signaling restraint or institutional discipline takes precedence for a leader over the need to mobilize public sentiment.

An examination of other recent conflicts, furthermore, suggests that this kind of induced caution around crisis signaling is not uncommon. During escalation between Rwanda and the Democratic Republic of Congo in 2025, for instance, neither President Kagame of Rwanda nor President Tshisekedi of the DRC prominently used Twitter/X for direct crisis messaging, even though both sides had used social media for both diplomatic signaling and domestic communication previously. And in clashes between Thailand and Cambodia in 2025, Thai Prime Minister Paetongtarn Shinawatra refrained from posting on social media during the crisis, relying instead primarily on press conferences, official statements, and diplomatic channels to communicate Thailand's stance.

The findings in this research note potentially offer one explanation for this strategic behavior: faced with very real dangers of crisis escalation, leaders might have incentives to avoid deploying technologies of affect through social media precisely because of their potential to stoke domestic audiences and given the very real dangers of inadvertent crisis escalation.

38. The crisis for the two countries was triggered on 14 February 2019 after a separatist militant rammed an explosive-laden vehicle into a convoy of Indian paramilitary soldiers in the Kashmiri district of Pulwama.

39. Narenda Modi, Twitter, 19 February 2019.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

This research note concludes with several suggestions for future research. First, if domestic audiences are indeed more likely to rally around a leader's social media messaging during an interstate crisis, and if domestic mobilization helps leaders signal credibility and resolve to external adversaries, it warrants asking whether this could inflate the incentives for leaders to take to social media to express escalatory intent or engage in "cheap talk" without actually planning on follow through. This concern has also been articulated by Greenhill, who notes that "rhetorical escalation can be a powerful, nonviolent method of simultaneously mobilizing support among audiences at home and signaling resolve to adversaries abroad … on the other hand, infusing rhetorical escalation with extra-factual information-laden messages is risky and can backfire, making both inadvertent and accidental escalation more likely."[40]

Second, while the vignettes employed in this study presented respondents with messages from the leader of their preferred political party, which might also explain the estimated effects on their subsequent sensitivity to nonpartisan elite cues, it could be argued that this constitutes a relatively easy test of domestic mobilization, and might not fully capture the dynamics of a broader "rally-'round-the-flag" effect. Would a leader's social media messaging be equally effective among outpartisans, and might outpartisans be more responsive to neutral, evaluative information even if it happened to be counter-attitudinal? This question is particularly important given the possibility that leaders facing domestic political challenges may have incentives to manufacture nationalist sentiment to consolidate support at home, dovetailing with findings from Barberá and Zeitzoff and Barberá and colleagues suggesting a connection between social unrest and a leader's social media adoption.[41]

Last, how representative is the social media content used in this research note of how leaders generally communicate during interstate crises? It is true that populist leaders may be especially prone to emotionally charged rhetoric online? But there are also leaders who might conceivably be more measured and calibrated on social media. This raises the question of scope conditions. First, I suggest that the arguments developed in this research note ought to apply most cleanly to contexts where leaders already maintain personalized social media accounts, and happen to find themselves in enduring interstate rivalries, where national identity frames are already primed and threat perceptions salient. The mechanisms described in this research note also require digitally connected domestic audiences (that is, contexts with some basic modicum of social media diffusion) and contexts where political attitudes are responsive to leader rhetoric. Moreover, I suggest that the effects described in this research note should likely be strongest in the early phases of crises where escalation pathways remain uncertain. When the risks of war become more immediate, as the case of the Pulwama–Balakot episode demonstrates, leaders may strategically avoid social media precisely to prevent domestic mobilization from constraining their policy options.

40. Greenhill 2024, 115.
41. Barberá and Zeitzoff 2017; Barberá et al. 2022.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

These arguments and ideas, I conclude, should hopefully encourage researchers to continue to unpack the many ways in which new communication technologies might shape leader behavior and domestic mobilization in the ever-evolving contexts of war and peace in our hyper-networked information age.

## Data Availability Statement

Replication files for this research note may be found at <https://doi.org/10.7910/DVN/WJL3R9>.

## Supplementary Material

Supplementary material for this research note is available at <https://doi.org/10.1017/S0020818326101362>.

## References

Amjadi, Eimon, and Richard S. John. 2025. Count on Me: Moral Language in Social Media and Policy Discourse during the Ukraine–Russia Conflict. *Data and Policy* 7:22.

Barberá, Pablo, and Thomas Zeitzoff. 2017. The New Public Address System: Why Do World Leaders Adopt Social Media? *International Studies Quarterly* 62 (1):121–30.

Barberá, Pablo, Anita R. Gohdes, Evgeniia Iakhnis, and Thomas Zeitzoff. 2022. Distract and Divert: How World Leaders Use Social Media During Contentious Politics. *The International Journal of Press/Politics* 29 (1):47–73.

Baum, Matthew A., and Tim J. Groeling. 2010. *War Stories: The Causes and Consequences of Public Views of War.* Princeton University Press.

Baum, Matthew A. 2004. Going Private: Public Opinion, Presidential Rhetoric, and the Domestic Politics of Audience Costs in US Foreign Policy Crises. *The Journal of Conflict Resolution* 48 (5):603–631.

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.

Baum, Matthew A., and Tim Groeling. 2008. New Media and the Polarization of American Political Discourse. *Political Communication* 25 (4):345–65.

Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys. *American Journal of Political Science* 58 (3):739–53.

Berinsky, Adam J., Michele F. Margolis, Michael W. Sances, and Christopher Warshaw. 2021. Using Screeners to Measure Respondent Attention on Self-Administered Surveys: Which Items and How Many? *Political Science Research and Methods* 9 (2):430–37.

Betsch, Tilmann. 2005. Preference Theory: An Affect-Based Approach to Recurrent Decision Making. In *The Routines of Decision Making*, edited by Tilmann Betsch and Steffen Haberstroh, 39–65. Lawrence Erlbaum Associates.

Bisaria, Ajay. 2024. *Anger Management: The Troubled Diplomatic Relationship Between India and Pakistan*. Aleph.

Boyd, Ryan L. 2017. Psychological Text Analysis in the Digital Humanities. In *Data Analytics in Digital Humanities*, edited by Shalin Hai-Jew, 161–89. Springer International.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

Boyd, Ryan L., Ashiqur Ashokkumar, Shahin Seraj, and James W. Pennebaker. 2022. The Development and Psychometric Properties of LIWC-22. University of Texas at Austin. Available at <https://www.liwc.app/help/psychometrics-manuals>.

Brutger, Ryan, Joshua D. Kertzer, Jonathan Renshon, Dustin Tingley, and Chagai M. Weiss. 2023. Abstraction and Detail in Experimental Design. *American Journal of Political Science* 67 (4):1007–1023.

Butt, Ahsan I. 2016. Street Power: Friday Prayers, Islamist Protests, and Islamization in Pakistan. *Politics and Religion* 9 (1):1–28.

Clary, Christopher, Sameer Lalwani, and Niloufer Siddiqui. 2021. Public Opinion and Crisis Behavior in a Nuclearized South Asia. *International Studies Quarterly* 65 (4):1064–1076.

Cohen, Stephen Philip. 2002. India, Pakistan and Kashmir. *Journal of Strategic Studies* 25 (4):32–60.

Coser, Lewis A. 1956. *The Functions of Social Conflict*. Simon and Schuster.

Croco, Sarah E. 2011. The Decider's Dilemma: Leader Culpability, War Outcomes, and Domestic Punishment. *The American Political Science Review* 105 (3):457–77.

Debs, Alexandre, and Jessica Chan Weiss. 2016. Circumstances, Domestic Audiences, and Reputational Incentives in International Crisis Bargaining. *Journal of Conflict Resolution* 60 (3):403–433.

Desposato, Scott, ed. 2016. *Ethics and Experiments: Problems and Solutions for Social Scientists and Policy Professionals*. Routledge.

Druckman, James N. 2001. On the Limits of Framing Effects: Who Can Frame? *Journal of Politics* 63 (4):1041–1066.

Eznack, Lucile. 2011. Crises as Signals of Strength: The Significance of Affect in Close Allies' Relationships. *Security Studies* 20 (2):238–65.

Eznack, Lucile. 2013. The Mood Was Grave: Affective Dispositions and States' Anger-Related Behaviour. *Contemporary Security Policy* 34 (3):552–80.

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

Fearon, James D. 1997. Signaling Foreign Policy Interests: Tying Hands Versus Sinking Costs. *Journal of Conflict Resolution* 41 (1):68–90.

Flavin, Patrick, and Michael T. Hartney. 2017. Racial Inequality in Democratic Accountability: Evidence from Retrospective Voting in Local Elections. *American Journal of Political Science* 61 (3):684–97.

Ganguly, Sumit. 2002. *Conflict Unending: India–Pakistan Tensions Since 1947*. Columbia University Press.

Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine.

Greenhill, Kelly. 2024. Of Wars and Rumors of Wars: Extra-Factual Information and (In)Advertent Escalation. In *Three Tweets to Midnight: Effects of the Global Information Ecosystem on the Risk of Nuclear Conflict*, 113–36. Hoover Institution Press.

Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Guisinger, Alexandra, and Alastair Smith. 2002. Honest Threats: The Interaction of Reputation and Political Institutions in International Crises. *Journal of Conflict Resolution* 46 (2):175–200.

Guisinger, Alexandra, and Elizabeth N. Saunders. 2017. Mapping the Boundaries of Elite Cues: How Elites Shape Mass Opinion Across International Issues. *International Studies Quarterly* 61 (2):425–41.

Haqqani, Husain. 2005. *Pakistan: Between Mosque and Military*. Carnegie Endowment for International Peace.

Harris, Benjamin Norwood, and Erik Lin-Greenberg. 2024. Cheap Tweets? Crisis Signaling in the Age of Twitter. *International Studies Quarterly* 68 (2).

Hayat, Noor, Muniba Fatima Zahra, and Ashbelah Shafaqat Ali. 2022. Politics in the Digital Age: Measuring the Impact of Twitter on Pakistani Youth. *Journal of Positive School Psychology* 6 (9):2267–78.

Hutchison, Emma, and Roland Bleiker. 2014. Theorizing Emotions in World Politics. *International Theory* 6 (3):491–514.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

Hymans, Jacques E.C. 2006. *The Psychology of Nuclear Proliferation: Identity, Emotions, and Foreign Policy*. Cambridge University Press.
Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly* 76 (3):405–431.
Jablanski, Danielle, Herbert S. Lin, and Harold A. Trinkunas. 2024. Retweets to Midnight: Assessing the Effects of the Information Ecosystem on Crisis Decision Making Between Nuclear Weapons States. In *Three Tweets to Midnight: Effects of the Global Information Ecosystem on the Risk of Nuclear Conflict*, 1–16. Hoover Institution Press.
Jost, Thomas, and Joshua D. Kertzer. 2024. Armies and Influence: Elite Experience and Public Opinion on Foreign Policy. *Journal of Conflict Resolution* 68 (9):1769–97.
Kenwick, Michael R., and Sarah Maxey. 2022. You and Whose Army? How Civilian Leaders Leverage the Military's Prestige to Shape Public Opinion. *The Journal of Politics* 84 (4):1963–78.
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., and Thomas Zeitzoff. 2017. A Bottom-Up Theory of Public Opinion About Foreign Policy. *American Journal of Political Science* 61 (3):543–58.
Kreps, Sarah. 2020. *Social Media and International Relations*. Cambridge University Press.
Levendusky, Matthew S., and Michael C. Horowitz. 2012. When Backing Down Is the Right Decision: Partisanship, New Information, and Audience Costs. *American Journal of Political Science* 56 (3):559–74.
Levy, Ro'ee. 2021. Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *American Economic Review* 111 (3):831–70.
McDermott, Rose. 2024. Psychological Underpinnings of Post-Truth in Political Beliefs. In *Three Tweets to Midnight: Effects of the Global Information Ecosystem on the Risk of Nuclear Conflict*, 17–38. Hoover Institution Press.
Mercer, Jonathan. 2014. Feeling Like a State: Social Emotion and Identity. *International Theory* 6 (3):515–35.
Mesquita, Rafael. 2024. What Do I Need to Say to Get Your Signature? Adding Draft Resolution Text to the UN General Assembly Sponsorship Dataset. *Research and Politics* 11 (4).
Milli, Smitha, Micah Carroll, Sashrika Pandey, and Yike Wang. 2023. Twitter's Algorithm: Amplifying Anger, Animosity, and Affective Polarization. Unpublished manuscript.
Mir, Asfandyar, and Niloufer Siddiqui. 2024. Nationalism, Status, and Conspiracy Theories: Evidence from Pakistan. *British Journal of Political Science* 54 (4):1159–79.
Morton, Rebecca B., and Joshua A. Tucker. 2014. Experiments, Journals, and Ethics. *Journal of Experimental Political Science* 1 (2):99–103.
Mueller, John E. 1973. *War, Presidents and Public Opinion*. Wiley.
Narang, Vipin, and Heather Williams. 2023. Thermonuclear Twitter? In *The Fragile Balance of Terror: Deterrence in the New Nuclear Age*, 63–89. Cornell University Press.
O'Loughlin, Ben. 2024. The Impact of the Information Ecosystem on Public Opinion during Nuclear Crises: Lifting the Lid on the Role of Identity Narratives. In *Three Tweets to Midnight: Effects of the Global Information Ecosystem on the Risk of Nuclear Conflict*, 179–92. Hoover Institution Press.
Pennebaker, James W. 2011. *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury.
Pornpitakpan, Chanthika. 2004. The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology* 34 (2):243–81.
Press, Daryl G., Scott D. Sagan, and Benjamin A. Valentino. 2013. Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons. *American Political Science Review* 107 (1):188–206.
Russett, Bruce. 1993. Can a Democratic Peace Be Built? *International Interactions* 18 (3):277–82.
Sagan, Scott D., and Benjamin A. Valentino. 2017. Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants. *International Security* 42 (1):41–79.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

Sagan, Scott D., and Benjamin A. Valentino. 2025. Atomic Arguments and Counter-Arguments: How Exposure to Conflicting Information Influences American Public Support for the Use of Nuclear Weapons. *International Studies Quarterly* 69 (1).
Sasley, Brent E. 2010. Affective Attachments and Foreign Policy: Israel and the 1993 Oslo Accords. *European Journal of International Relations* 16 (4):687–709.
Schelling, Thomas C. 1966. *Arms and Influence*. Yale University Press.
Schultz, Kenneth A. 1998. Domestic Opposition and Signaling in International Crises. *The American Political Science Review* 92 (4):829–44.
Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy. Cambridge Studies in International Relations*. Cambridge University Press.
Schultz, Kenneth A. 2017. Perils of Polarization for US Foreign Policy. *The Washington Quarterly* 40 (4):7–28.
Simmons, N. 2017. Axial Coding. In *The SAGE Encyclopedia of Communication Research Methods.* Vol. 4, 80–82. SAGE.
Slantchev, Branislav L. 2006. Politicians, the Media, and Domestic Audience Costs. *International Studies Quarterly* 50 (2):445–77.
Stroud, Natalie. 2010. Polarization and Partisan Selective Exposure. *Journal of Communication* 60 (30):556–76.
Tausczik, Y.R., and J.W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29 (1):24–54.
Thachil, Tariq. 2017. Do Rural Migrants Divide Ethnically in the City? Evidence from an Ethnographic Experiment in India. *American Journal of Political Science* 61 (4):908–926.
Thachil, Tariq. 2018. Improving Surveys Through Ethnography: Insights from India's Urban Periphery. *Studies in Comparative International Development* 53 (3):281–99.
Tomz, Michael. 2007. Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization* 61 (4):821–40.
Tomz, Michael R., and Jessica L.P. Weeks. 2013. Public Opinion and the Democratic Peace. *American Political Science Review* 107 (4):849–65.
Tomz, Michael R., and Jessica L.P. Weeks. 2020. Human Rights and Public Support for War. *Journal of Politics* 82 (2):500–517.
Wayne, Carly N. 2023. Terrified or Enraged? Emotional Microfoundations of Public Counterterror Attitudes. *International Organization* 77 (4):824–47.
Weeks, Jessica L. 2008. Autocratic Audience Costs: Regime Type and Signaling Resolve. *International Organization* 62 (1):35–64.
Weiss, Jessica Chen. 2014. *Powerful Patriots: Nationalist Protest in China's Foreign Relations*. Oxford University Press.
Weiss, Jessica Chen, and Allan Dafoe. 2019. Authoritarian Audiences, Rhetoric, and Propaganda in International Crises: Evidence from China. *International Studies Quarterly* 63 (4):963–73.

## Author

**Fahd Humayun** is Assistant Professor of Political Science at Tufts University. He can be reached at fahd.humayun@tufts.edu.

## Acknowledgments

The author thanks Alexandre Debs, Babak Bahador, Erik Lin-Greenberg, Josh Kertzer, Steven Wilkinson, participants at the 2023 *IO* Café in Montreal as well as three anonymous reviewers for their constructive engagement. Thank you as well to Zehra Wasif at the Lahore University of Management Sciences for helpful research assistance. IRB Protocol ID: 2000032761.

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press

## Funding

This research note was funded through the support of the Institution for Social and Policy Studies (ISPS) at Yale University.

## Key Words

Domestic mobilization; crisis escalation; leaders; social media; nationalism; South Asia

Date received: August 07, 2025; Date revised: March 23, 2026; Date accepted: March 23, 2026

https://doi.org/10.1017/S0020818326101362 Published online by Cambridge University Press