Proceedings of the Twentieth International AAAI Conference on Web and Social Media (ICWSM 2026)

From Attention to Dialogue: Does Audience Engagement Reinforce Constructive Cross-Party Communication?

Ahana Biswas, Yu-Ru Lin
University of Pittsburgh
ahana.biswas@pitt.edu, yurulin@pitt.edu

Abstract

While existing works have emphasized how elites shape mass opinion, we ask whether the reverse also holds: do audience reactions on social media actively shape elite behavior? We examine this question through the lens of cross-partisan interactions (CPIs), which can either foster deliberation or deepen polarization. Using a dataset of over 1.1 million cross-party retweets, replies, and mentions between U.S. state legislators and their audiences on Twitter/X (2020–2021), we first establish baseline patterns of engagement: Democrats gain modest engagement in replies and mentions, while Republicans often face penalties in direct cross-party interactions. Building on this, we show that audience engagement produces a feedback loop that conditions future elite behavior. Following highly visible CPIs, legislators are not only more likely to engage again in cross-talk, but also shift their rhetorical strategies. Engagement consistently promotes causal reasoning, subjective language, and positive-emotion framing in subsequent CPIs. These findings suggest a positive association between audience engagement and constructive cross-party discourse among elites, challenging overly simplified interpretations in the literature that emphasize social media as a primary driver of rising or falling polarization.

1 Introduction

Cross-partisan interactions (CPIs)—instances of engagement across party lines—are often regarded as vital to democratic discourse. Such exchanges can foster deliberation and reduce stereotyping across divides (Habermas 1989; Pettigrew and Tropp 2006). Yet they can also trigger backlash, amplify hostility, and heighten affective polarization (Mutz 2002; Bail et al. 2018). This tension raises a fundamental question: do CPIs help depolarize, or do they reinforce partisan divides?

This question is especially pressing in the case of political elites, whose communication carries disproportionate weight in shaping public opinion. Elite cue theory holds that political leaders provide informational shortcuts that shape how citizens interpret and respond to political issues (Zaller 1992). More recently, research on online influencers highlights how highly visible figures strategically manage their self-presentation and audience engagement in order to capture and sustain attention (Tufekci 2013). Legislators today occupy both roles: as elites, their communication has downstream consequences for public opinion, and as visible figures on social media, they are embedded in dynamics where engagement is currency. This dual role raises an important question: *In the attention economy, do audiences shape elites' behavior by rewarding certain cross-party interactions while discouraging others?*

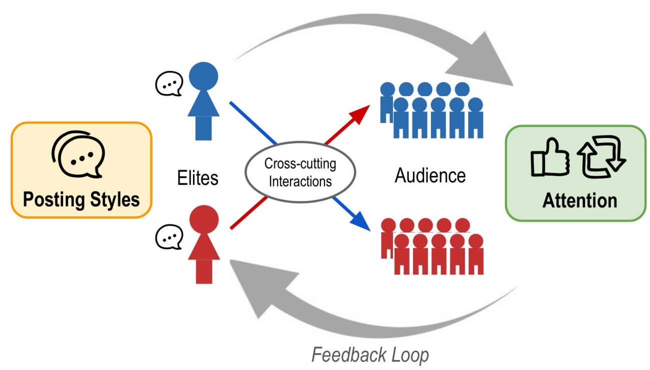

Figure 1: **Study Design.** We analyze how CPIs shape the online engagement of political elites and how engagement feeds back into future behavior. We test whether CPIs boost or reduce engagement, and which posting styles drive these effects (RQ1), and how engagement shapes both the frequency (RQ2) and style (RQ3) of future CPI.

To investigate this possibility, we deliberately narrow our scope. First, we focus on political elites rather than general users, since their communication is uniquely positioned to shape democratic discourse. Legislators, while not influencers in the conventional sense, serve as a critical proxy for elites whose online behavior has both symbolic and strategic weight. Second, we examine cross-partisan posts, as these interactions are especially consequential for polarization and depolarization (Nyhan and Reifler 2010; Garimella et al. 2018). Third, we concentrate on audience[1] engagement, since engagement is the primary way in which publics

[1]We use the term *audience* to refer to users who engage with a legislator's posts. For instance, references to "Republican (Democratic) audiences" denote the audiences of Republican (Democratic) legislators.

reward or penalize elites online and prior research suggests it conditions future communicative behavior (Slater 2007). Finally, we study not only whether elites engage across party lines, but also how they do so—capturing both the frequency and stylistic choices of CPIs as outcomes that may shift in response to prior engagement[2].

To conceptualize these dynamics, we adopt the stimulus–organism–response (SOR) model (Jacoby 2002). In this framework, audience engagement functions as the stimulus, elites' interpretation of these signals represents the organismal process, and subsequent posting frequency and style are the responses. This lens allows us to move beyond static assessments of engagement to test whether attention feeds back into elites' future communication strategies. Our focus thus shifts from audience reactions alone to the feedback loops that connect audience engagement and elite behavior—an aspect of bipartisan communication that has received limited attention in prior research. In this work, we focus on the positive direction of this loop, asking how increases in the engagement of CPIs shape the likelihood and style of future cross-talk. We hypothesize that positive reinforcement dominates because platforms foreground visible signals of success (e.g., *likes, replies, shares*), giving elites stronger incentives to repeat attention-getting behaviors.

Our study leverages over 1.2M interactions between U.S. state legislators and their audiences on Twitter/X during 2020–2021. We organize our analysis around four RQs:

RQ1: How does cross-partisan communication gain or lose engagement in the attention economy?

- **RQ1a:** To what extent do cross-party interactions, compared to intra-party ones, shape the engagement of elites?
- **RQ1b:** How does the nature of cross-talk (e.g., tone, framing, topics) affect their engagement?

RQ2: How does the engagement of past cross-party interactions affect how often legislators engage in future cross-partisan communication?

RQ3: To what extent does the engagement of past cross-party interactions affect the rhetorical or stylistic choices legislators make in future cross-partisan communication?

RQ1a establishes the baseline engagement effects of CPIs, RQ1b examines which posting styles explain these effects, and together they set the stage for RQ2 and RQ3, which test whether audience responses feed back into the frequency and style of elites' future cross-partisan behavior (Fig. 1[3]).

By framing CPIs as part of a dynamic feedback loop, this study extends prior work on cross-cutting exposure and partisan discourse by centering elite behavioral change. Our findings offer new insight into whether social media incentivizes constructive cross-party dialogue or reinforces divisive, attention-maximizing strategies—with direct implications for understanding polarization and the health of democratic communication.

Findings. Our findings reveal marked asymmetries. Republican legislators see lower engagement when engaging via direct modes of communication like replies and mentions, whereas Democrats experience modest engagement gains (+4.5% and +8%). Content also matters. Incivility in cross-party replies increases Republican engagement by 18%, while uncivil mentions reduce engagement for both parties. Importantly, higher engagement with CPIs is linked to increased bipartisan communication in subsequent weeks: both parties increase mentions (by 2.2% for Republicans, 1.3% for Democrats) following highly visible cross-party posts. Beyond frequency, we also find that engagement conditions how legislators communicate in future CPIs. For replies, engagement is positively associated with hedged (+0.9%) and subjective (+1.2%) framing for Democrats, and subjective and causal framing (both by 0.9%) for Republicans. For mentions, Democrats lean into subjective and positive-emotion expression by 0.7%), while Republicans reduce URLs (-1.5%) but amplify hedging, causality, positive, and issue-centered mentions by around 1%. Taken together, these results provide evidence of a engagement-driven feedback loop in which audience responses shape not only whether but also how elites engage across party lines.

Implications. Prior work on digital political communication often highlights how audience engagement reinforces polarizing behaviors, such as antagonism or incivility, raising concerns about the incentives built into attention-driven platforms (Mutz, 2002; Bail et al., 2018; Cinelli et al., 2021). Our findings extend this literature by showing that engagement is associated not only with amplification or polarization, but also with subsequent shifts in communicative style. Specifically, higher engagement in CPIs is positively associated with increased use of subjective, causal, and topical language in later replies and mentions. These stylistic features are commonly linked to explanation, contextualization, and issue-focused discussion, which are often considered components of deliberative exchange. These results suggest that audience engagement may coincide with communicative adjustments in cross-party interaction. This points to a more nuanced role for engagement in democratic communication: engagement may be aligned with discourse strategies that support more substantive interaction across partisan lines, rather than uniformly reinforcing polarization.

2 Related Work

Exposure to Cross-Cutting Discussions Scholars have long debated whether cross-cutting exposure strengthens or weakens democracy. Some view it as fostering deliberation and reducing stereotyping across divides (Habermas 1989; Barber 2003; Pettigrew and Tropp 2006), while others argue it discourages participation or exacerbates polarization (Mutz 2002; Taber and Lodge 2006; Nyhan and Reifler 2010). Empirical work documents both dynamics: dis-

[2]Throughout the paper, we use the term *engagement* to denote a user-normalized measure of audience response, defined relative to each legislator's own recent posting history. Specifically, engagement captures whether a given post receives more or fewer responses than that legislator's typical recent posts, rather than measuring absolute exposure or impressions (see Section 3.3).

[3]Fig. 1 icons by Andrew Aprilio Raharjo (CC BY 3.0)

agreeable exposure can mobilize political participation (Lu et al. 2016; Kim and Chen 2016; Min and Wohn 2018), but can also provoke backfire effects (Nyhan and Reifler 2010; Taber and Lodge 2006). These effects are often asymmetric across the ideological spectrum, as liberals and conservatives prioritize distinct values (Graham et al. 2009; Jost et al. 2007; Grossmann and Hopkins 2016). Computational studies extend this debate to online contexts: heterogeneous discussions do occur (An et al. 2019; Wu and Resnick 2021), yet they can also intensify divides (Bail et al. 2018). Moreover, bipartisan actors are often less central and attract lower engagement (Garimella et al. 2018). Taken together, this body of work underscores that engagement with CPIs is far from uniform. Some contexts promote deliberation and broaden participation, while others intensify polarization. These mixed findings raise a critical question about how engagement dynamics contribute to cross-talk online.

Nature and Style of Political Cross-Talk Cross-party communication is a recurring but contested feature of online discourse (Wu and Resnick 2021; Morales et al. 2021). Some interactions aim at genuine deliberation, while others resemble trolling or provocation (Phillips 2015; Flores-Saviaga et al. 2018; Hua et al. 2020). Linguistic and emotional patterns also vary: users adapt their style when addressing out-partisans compared to co-partisans (An et al. 2019), and CPIs are often more negative or emotional (Burrell 2020; Wu and Resnick 2021; Marchal 2022). Broader studies show partisan divides in framing during political crises (Demszky et al. 2019), yet linguistic cues such as positivity or reduced subjectivity can foster engagement across ideological lines (Saveski et al. 2022). These highlight the importance of studying the role of linguistic features and topics in driving content engagement, as they can influence how CPIs resonate across ideological divides and either foster more inclusive or divisive political discourse.

Feedback Loops and Elite Adaptation Political science has long emphasized that elites are responsive to public signals and strategically adapt their communication (Downs 1957; Carrubba 2001; Gabel and Scheve 2007; Santoro et al. 2021). Traditionally, responsiveness was inferred from elections, opinion polls, or media coverage. On social media, elites instead receive instantaneous feedback in the form of engagement metrics that signal audience approval. Computational studies highlight how such feedback loops shape behavior in digital environments: scientists adapt their messaging after viral attention (Hasan et al. 2022), newsrooms shift coverage toward high-engagement stories (Caplan and Boyd 2016), and creators adjust strategies to algorithmic incentives (Bishop 2019). Feedback also shapes political behavior, as algorithmic amplification can reinforce user actions including misinformation sharing (Cinelli et al. 2021; Hanley and Durumeric 2023). Communication research has long theorized reinforcing spirals between exposure and expression (Slater 2007), but large-scale empirical evidence of feedback-driven adaptation in elite cross-party discourse remains limited. Whether audience engagement systematically shapes elites' CPIs is an open question.

Present Work. Despite extensive work on cross-cutting exposure, cross-talk styles, and elite responsiveness, few studies examine whether audience engagement feeds back into elite cross-party behavior. Prior work shows that bipartisan actors often struggle for engagement (Garimella et al. 2018), and that audiences are more responsive to out-party criticism than in-party praise (Yu et al. 2024). However, these studies do not examine if engagement systematically alters elite strategies over time. We address this gap by analyzing over one million CPIs between U.S. state legislators and their audiences on Twitter/X. Our study proceeds in two stages. First, we examine how cross-party interactions affect elite engagement and how posting styles shape engagement outcomes **(RQ1)**. Second, we investigate whether engagement feeds back into elite behavior by influencing the frequency **(RQ2)** and posting style **(RQ3)** of future CPIs. By linking audience engagement to elite adaptation, we position CPIs as embedded in a feedback system, where engagement acts as a reinforcement signal.

3 Study Design

Our study proceeds in two stages to examine how engagement shapes cross-partisan interactions. First, we analyze how different forms of cross-party engagement and associated rhetorical and stylistic choices resonate with audiences in the attention economy. Building on these patterns of audience response, we then examine how engagement signals feedback into elites' subsequent communicative behavior. The following sections describe the data, variables, and modeling strategy used in our analysis.

3.1 Dataset

We use the Twitter dataset provided by Biswas et al. (2025), which contains the complete posting history of all U.S. state representatives and senators who held office at any point during 2020-2021. This dataset includes approximately 4 million tweets posted by 3,568 (38.8%) U.S. state legislators, of which 53.4% are Democrats, responsible for 71.6% of the posts, and 45.3% are Republicans, contributing 28.3% of the posts[4]. Out of the these, around 2.3M (57.5%) tweets had atleast one interaction (i.e., either replied to/mentioned/retweeted another user)[5]. Following prior work (Biswas et al. 2025), we measure the engagement with each post by summing the number of likes, retweets, replies, and quotes.

Additionally, we collect all available follower accounts from a 2021 data snapshot, which allows us to construct a follower network of the legislators. This network consists of approximately 3.8 million connections and over 1.3 million unique follower accounts. The set of followers (U_F) and the set of legislators (U_L) together form the set of users ($U = U_F \cup U_L$) considered in this study.

[4]The number of Independent legislators (N=29) is too small for meaningful analysis and has thus been excluded from the study.

[5]This accounts for interactions with 1.7M unique users. Around 73% of these interactions originate from Democrat legislators.

Cross-cutting Interactions. To explore the relationship between political interactions and public engagement, we define "cross-cutting" interactions as communication that occurs across partisan lines. Specifically, we define a legislator's post as *cross-cutting* if it mentions, replies to, or retweets a user in U—whether a legislator or follower—whose political leaning opposes their own[6]. While the political affiliations of legislators are readily available, the leanings of their followers are not. To infer the political leaning of a follower $f \in U_F$, let D_f and R_f represent the number of Democratic and Republican legislators, respectively, that user f follows. We include only those followers who follow at least three legislators to ensure a more reliable estimation of their leaning.

The leaning score l_f of each user f is calculated as the normalized difference between the number of Democratic (D_f) and Republican (R_f) legislators followed by the user. We refer to this metric as the "attention leaning" since it reflects the partisan distribution of legislator accounts followed. Specifically, the score is given by:

$$l_f = \frac{D_f - R_f}{D_f + R_f} \tag{1}$$

This results in a score ranging from -1 to 1, where a score closer to 1 indicates a stronger alignment with Democrats, and a score closer to -1 indicates a stronger alignment with Republicans. Users who follow an equal number of Democrat and Republican legislators will have a score of 0, indicating no clear partisan preference. Figure 2 shows the distribution of attention leaning scores for various interaction types [7]. This distribution highlights the political polarization in the interactions, where Democrat legislators predominantly interact with users whose scores are closer to 1, while Republicans tend to interact with users whose scores are closer to -1, suggesting that legislators' interactions on Twitter are heavily aligned with users who share their political affiliation.

Based on the distribution of attention leaning scores and manual annotation of a sample of their Twitter accounts (see Appendix for details), we set the threshold for identifying Democrat-leaning followers at a score of 0.71 or higher and Republican-leaning followers at -0.69 or lower[8]. Interactions with users from the same political leaning are considered as *intra-partisan* posting. Using our method we are able to identify the partisan leaning of users interacted with for around 1.17M (50.8%) posts by 3,287 (92.1%)[9] legislators which is the final dataset used in our study. Among these around 82K (7%) posts have atleast one cross-cutting interaction[10].

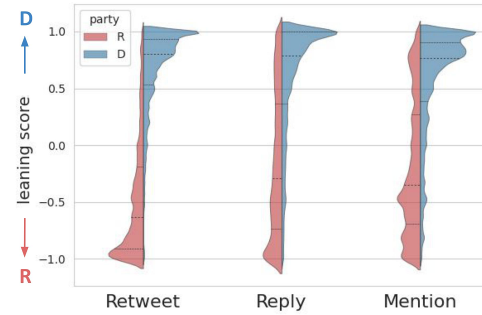

Figure 2: **Distribution of attention scores.** Legislators from both parties primarily engage with users having the same political leaning across all interaction types. Republicans tend to engage in more cross-talks than Democrats.

		N	Engagement
Retweet	R $\rightarrow$ D	13,520 (10.9)	29.0 (6.4)
	D $\rightarrow$ R	8,595 (60.9)	3291.5 (0.007)
Reply	R $\rightarrow$ D	11,133 (2.1)	1.0 (1.0)
	D $\rightarrow$ R	7,040 (16.8)	1.0 (2.0)
Mention	R $\rightarrow$ D	35,709 (2.2)	5.0 (1.2)
	D $\rightarrow$ R	29,168 (17.9)	8.0 (0.75)

Table 1: Volume of cross-cutting posts by each party and their median engagement. The numbers in the brackets denote the ratio of not cross-cutting to cross-cutting posts.

Table 1 shows the volume and Figure 8 in Appendix shows the rate of cross-cutting and intra-partisan posts and their engagement. Republican legislators have 5 to 6 times more cross-partisan interaction rates than their Democratic counterparts (Fig. 8A). For instance, among all replies from Republican legislators, 32% are cross-partisan, compared to only 5.6% of replies from Democrats. The engagement metric for retweets returned by the Twitter API inherently reflect the popularity of the original post (see Appendix), therefore we only report results for mention and reply in the paper. Results for retweet are provided in the Appendix. The typical engagements on cross-cutting posts are usually lower than intra-partisan ones for Republicans and in the case of replies for Democrats. The engagement is notably higher when Democrats retweet users having Republican leaning (Fig. 8B).

[6]Interactions are included in our dataset if a legislator engages with a user's content (e.g., retweets, mentions, or replies) and the user's political leaning can be inferred from the follower network, even if the user does not follow the legislator.

[7]Mentions in our analysis do not account for replies, i.e., an account is considered "mentioned" only if it explicitly appears in the text body of the post as a mention.

[8]This yields over 53.6K (3% of all users interacted with) unique users with either Republican or Democrat leaning. These users appear to be highly politically attentive: they follow multiple legislators and account for a disproportionate share (approximately 50%) of elite-audience interactions.

[9]On average, we are able to estimate 55% interaction directions (i.e., cross-party or intra-party) for each legislator overall (49% and 56% for Republicans and Democrats respectively).

[10]Notably, a single tweet can contain multiple interactions.

3.2 Individuals' Attributes

We characterize legislators by their activities on the platform and individual-level characteristics. To measure platform activity and influence, we include features such as post count, follower count, and in-degree centrality within the follower network, serving as proxies for their network influence (Hasan et al. 2022).

The individual-level attributes include socio-demographic features such as state, gender (Men vs. Women), ethnicity[11] (White vs. Non-White), and ideology scores (Biswas et al. 2025). The dataset consists of approximately 68.2% men and 62.2% White legislators. We leverage the ideology scores constructed by Shor and McCarty (2011) where the scores range between -2.3 (extremely liberal) to +2.3 (extremely conservative).

In addition to the socio-demographics, we include two other individual-level attributes, (i) central (C) vs. local (L) and (ii) audience partisanship.

Central (C) vs. Local (L). We categorize legislators as central, local, or balanced based on the geographic diversity of their audiences. Central legislators, with broader cross-state followings, can shape national discourse by engaging conversations that span diverse viewpoints (Mutz 2001; Jun 2014). Their CPI may help disrupt echo chambers by connecting opposing perspectives. Local legislators, by contrast, may concentrate on state-specific audiences, tailoring messages to regional concerns and making cross-partisan communication more salient at the local level. Audience locality is measured as the fraction of followers from a legislator's home state; higher scores indicate more localized audiences (See Appendix). Legislators in the top and bottom 30 percentiles are classified as local and central, respectively.

Audience Partisanship. CPIs can be effective in bridging political divides or may have backfire effects depending on the partisan composition of a legislator's audience. Audience partisanship captures the extent to which a legislator's audience leans toward one political party, which can shape how cross-partisan interactions are received. We operationalize audience partisanship using the attention-leaning score of followers (Garimella and Weber 2017). Specifically, the audience partisanship score of a legislator is measured as the median leaning score of users who engage with that legislator's posts. This measure does not capture ideological homogeneity or affective polarization within the audience, but rather provides a summary measure of partisan orientation that is relevant for understanding how CPIs are evaluated by different party audiences.

3.3 Measuring Engagement

Our measure of engagement, termed as *overperforming* score, is drawn from prior work by Biswas et al. (2025)[12]. The overperforming score for post i is, $y_{ui} = v_{ui}/(b_u + b_0)$, where v_{ui} is the engagement received on post i, b_u is the median engagement for legislator u's posts on the platform in previous m-days and a threshold (b_0) for the minimum number of engagements on a post to be considered as overperforming. We choose the $b_0 = 10$ and $m = 14$ as suggested by Biswas et al. (2025) (See Appendix) . This metric captures engagement relative to a legislator's own recent baseline, thereby normalizing for stable differences across legislators (e.g., audience size) and slowly varying platform-level factors that affect engagement broadly. While we do not explicitly model platform ranking algorithms or exposure, the baseline normalization reduces sensitivity to global shifts in platform dynamics that influence engagement across posts within the same time window.

4 Methods

In this section, we outline the characteristics of legislators' posting styles and discussing their relevance. Finally, we detail the methods used to address our research questions.

4.1 Characterizing Posting Styles

We select posting styles commonly studied in political communication. Broadly, we study two categories: rhetorical styles, which capture how legislators frame arguments and reasoning (e.g., causality terms, generalization, hedging, topic indicators, URL usage), and stylistic features, which reflect tone and affect (e.g., incivility, positive or negative emotion, subjectivity). This allows us to examine how different dimensions of communication map onto audience engagement and future cross-party interactions.

Linguistic Markers. We use four linguistic cues—*emotion*, *subjectivity*, *generalization*, and *argumentation*—to characterize the legislators' posting styles. These features capture distinct aspects of political communication—emotion highlights the intensity of posts, subjectivity reflects the degree of personal opinion, generalization assesses accessibility and inclusiveness, and argumentation measures logical reasoning and persuasive quality. Below we describe our reasoning and the choice of measures for these markers.

Emotion. Prior work shows that emotionally charged content elicits stronger audience responses and engagement (Mutz 2007). Emotional language can enhance the perceived sincerity of a message, shaping both attention and public opinion. In polarized contexts, such appeals may reinforce or challenge existing beliefs, making emotion a key marker for CPI. We measure emotions using positive and negative word lists from Linguistic Inquiry and Word Count (LIWC)[13]. Among negative emotions, we focus on anger and anxiety, which are especially influential in political discourse (De Castella and McGarty 2011; Ryan 2012).

Subjectivity. Subjectivity refers to the expression of opinions and evaluations (Wiebe et al. 2000). In cross-party interactions, subjective language can increase engagement

[11] Ethnicity and gender are mapped using Ballotpedia. Binary genders used due to lack of sufficient data on non-binary genders.

[12] This score is inspired by the metric used at Meta CrowdTangle and generalized to be suitable for other platforms like Twitter. The engagement components (likes, shares, replies, and quotes) used to get *engagement* are similarly used on Twitter (https://github.com/twitter/the-algorithm?tab=readme-ov-file) to calculate a post's expected engagement.

[13] https://www.liwc.app/

by presenting relatable or persuasive viewpoints, but may also polarize by emphasizing partisan perspectives. We measure subjectivity using OpinionFinder's Subjectivity lexicon (Wilson et al. 2005).

Generalization. Generalizations can make cross-party messages more accessible to broad audiences but also risk stereotyping and reducing nuance, potentially reinforcing divides. Analyzing generalizations helps assess whether such communication balances simplicity with accuracy and inclusiveness in shaping engagement. We measure them using MPQA's[14]generalization lexicon.

Argumentation. Argumentation conveys reasoning and persuasion, making it central to constructive CPIs. Well-structured arguments can foster productive communication and engage diverse audiences. We capture argumentation through *hedges* and *causation*[15]. Hedge words signal uncertainty and can soften confrontational tones, fostering receptiveness across divides, and are measured using the lexicon from Islam et al. (2020). Causation terms clarify connections between events and provide logical structure, enhancing persuasiveness in cross-party dialogue; we use word lists from MPQA and LIWC to measure them.

We evaluate the quality of the lexicon-based markers by manually labeling around 40 samples per linguistic feature (Appendix). The precision of certain markers differs across the parties, possibly due to the distinct communication styles employed by the parties. To mitigate this variation, our study design incorporates a comparative approach by evaluating cross-party interactions relative to intra-party interactions as described in later sections. This design helps to isolate the effects of cross-party communication while minimizing biases arising from party-specific linguistic styles. Additionally, the prevalence of each posting style and representative examples for both parties are reported in Appendix Table 7.

Harmful Content. *Harmful* content, including uncivil language, and untrustworthy or non-credible information can play a significant role in shaping the dynamics of cross-party discussions, especially in politically charged environments (Goovaerts et al. 2020; Hanley and Durumeric 2023).

Incivility. Incivility refers to disrespectful, hostile, or derogatory language. In cross-party interactions, it can escalate tensions, polarize audiences, and discourage constructive communication. We assess incivility using toxicity scores, a common practice in prior work (Frimer et al. 2023; Kim et al. 2021). Following (Biswas et al. 2025), we use Detoxify to generate toxicity scores and apply a threshold of 0.82[16], classifying posts above this cutoff as uncivil. Around 2.3% of Democratic CPIs (0.6% intra-partisan) and 1.1% of Republican CPIs (0.7% intra-partisan) are uncivil.

Low-credibility. Low-credibility content in cross-party discussions can undermine debate quality, reduce trust, and mislead the public. When political elites share such information, they risk perpetuating false narratives and weakening democratic processes. We detect low-credibility content using labels from Tai et al. (2023), which rely on Media Bias/-Fact Check (MBFC) credibility ratings of URL domains[17], a common practice in prior work (Lasser et al. 2022). Around 0.3% of CPIs from both Republicans (2.1% intra-partisan) and Democrats (negligible intra-partisan) contain low-credibility information (see Appendix).

Topics. Topics frame the content of CPI and shape both the nature of interactions and public engagement. We classify posts using a keyword-based approach followed by semantic clustering, covering salient issues in U.S. politics such as *BLM, COVID-19, rights, immigration, gun control, climate, abortion*, and the *Capitol riots* (see Appendix). Posts without relevant keywords are labeled "other," while those matching multiple categories are assigned to all relevant topics to capture complex, multi-faceted discussions.

External Information. Posts containing URLs may provide additional information, attracting users with diverse perspectives and encouraging engagement across viewpoints. We consider the inclusion of a URL[18] as a feature of posting style that could influence the engagement of CPIs.

Differences in Cross-party vs. Intra-party Posting Styles. We compare posting styles[19] and legislators' attributes across cross- and intra-party interactions using Mann–Whitney U tests with Bonferroni correction. Effect sizes (rank-biserial correlations) are reported in Appendix Fig.9 (confidence intervals in Fig.10).

Clear differences emerge. Negative emotions such as anger, anxiety, and incivility appear more often in deliberative forms of cross-talk (replies and mentions), echoing prior findings that cross-partisan communication tends to be more toxic or antagonistic (Burrell 2020; Wu and Resnick 2021). Legislator attributes also matter: among Democrats, central legislators are more likely to engage across party lines, while among Republicans, local legislators take the lead. In both parties, legislators with polarized audiences engage less in all forms of cross-talk, and those at the ideological extremes are least likely to participate, consistent with theories of selective engagement (Heatherly et al. 2017).

These results suggest that bipartisan communication is shaped not only by rhetorical styles but also by legislators' representational contexts and audience structures. These variations can strongly influence the engagement of CPIs and are incorporated into our analyses in the following sections.

[14]https://mpqa.cs.pitt.edu/lexicons/arg_lexicon/

[15]Argumentation mining is not feasible due to short text length.

[16]Biswas et al. (2025) identified 0.82 as the optimal threshold for this dataset based on manual labeling

[17]Domain-level information was extracted from the expanded URL metadata when available, or from the visible domain portion of shortened or truncated URLs, which is sufficient for identifying source domains.

[18]We do not examine URL types (e.g., finance, sports) in detail, which could offer additional insights but is limited by the challenge of recovering full URLs for all posts.

[19]All posting styles are binary, i.e., 1 if present. For linguistic markers, the marker is considered present if the lexicons appear at least once.

4.2 Measuring the Effect of Cross-cutting Interactions on Engagement (RQ1)

We estimate the observed causal effect of cross-cutting interactions on engagement using a two-fold approach: matching and regression analysis.

Matching. CPIs systematically differ from intra-party ones in ways that may also affect engagement. To reduce this bias, we match each CPI to an intra-party counterpart within the same interaction type (retweet, reply, mention) and party. This addresses potential confounders such as topical focus (e.g., gun control), rhetoric strategies, legislator attributes (Section 4.1), and timing (e.g., election periods). Posts are represented using RoBERTa embeddings (Liu et al. 2019), concatenated with legislator attributes and timing; for low-credibility content, URL headlines are also included. Posts shorter than 10 words are excluded. Following prior work (Hasan et al. 2022; Sahly et al. 2019; Biswas et al. 2025), these covariates capture key textual and contextual factors. We then perform 1:1 K-Nearest Neighbors matching, which achieves covariate balance (Appendix Fig. 11).

Regression Analysis (RQ1a). Using the matched posts, we estimate the (observed) effect of cross-cutting interactions using a linear mixed effects model,

$$y_{ui} \sim \alpha_0 + \alpha_1 \Theta_{ui} + \vec{\alpha}_X X_u + r_u + d_i \tag{2}$$

where y_{ui} is the overperforming score of legislator u's post i and Θ_{ui} is a binary variable indicating whether the post is cross-cutting (i.e., $\Theta = 1$). Therefore, α_1 captures the effect of cross-cutting interactions on the engagement of legislators' posts. X_u is the vector of individual attributes described in Section 3.2 used to control for confounding at the user-level. We incorporate random effects on the user and post timing (i.e., day count) denoted by r_u and d_i respectively—to control for unobserved heterogeneity, i.e., factors that vary across users or time periods but are not explicitly included in our model. This is done to reduce bias and improve the robustness of the estimated effects.

Effect of Posting Styles (RQ1b). We analyze how posting styles (see Section 4.1) are associated with the engagement of cross-cutting posts. Unlike RQ1, which examines a single treatment variable, we decompose the variable from 2 into two parts: the effects of posting styles and other uncaptured factors. We use the following regression analysis,

$$y_{ui} \sim \alpha_0 + \alpha_{L^-} \Theta_{ui} + \vec{\alpha}_L L_{ui} \Theta_{ui} + \vec{\alpha}_X X_u + r_u + d_i \tag{3}$$

where L_{ui} is the vector of posting styles of legislator u's post i and the coefficient vector $\vec{\alpha}_L$ represents the effect of each posting style on cross-talk engagement. The coefficient α_{L^-} accounts for the treatment effects not captured by the posting style variables. The other variables are same as described in equation 2. We similarly use a linear mixed effects model to estimate equation 3.

4.3 Assessing the Link between Engagement and Future Cross-party Interactions (RQ2)

In RQ2, we investigate how the engagement on cross-cutting interactions is linked to the likelihood of future cross-party talks from political elites. More specifically, we estimate how changes in engagement are correlated with the rate of cross-party interactions in the subsequent week. This is done using the following regression analysis:

$$\rho_{u,t+1} \sim \beta_0 + \beta_1 \delta_{u,t} + \vec{\beta}_X X_u + r_u + w_t + \rho_{u,t} \tag{4}$$

Here, $\rho_{u,t}$ and $\rho_{u,t+1}$ represent the fraction of cross-party posts made by legislator u in weeks t and $t+1$, respectively. The term $\delta_{u,t} = \text{frac}(y_{u,t|c} > 1) - \text{frac}(y_{u,t|nc} > 1)$ measures the difference between the fraction of overperforming cross-cutting posts and non-cross-cutting posts (nc) in week t. A post is considered "overperforming" if its engagement metrics exceed a threshold of 1. Non-cross-cutting (nc) posts refer to intra-partisan posts or posts with no interactions or interactions with users whose leaning is not known (i.e., all posts other than cross-cutting ones). The coefficient β_1 represents the effect of changes in engagement on the future rate of cross-party interactions. To account for variability across individuals and time, we include random effects for users (r_u) and weeks (w_t), capturing individual-specific and temporal fluctuations, respectively. The term $\rho_{u,t}$ accounts for the autoregressive effect, where the current cross-cutting posting rate influences future interactions.

4.4 Engagement and the Posting Style of Future Cross-party Interactions (RQ3)

In RQ3, we examine whether the engagement gained from cross-party interactions bis associated with the stylistic choices elites adopt in subsequent cross-party communication. We test whether legislators use of particular posting styles—such as incivility, emotional tone, or topical framing changes depending on the relative engagement of their cross-partisan posts. We model this relationship as follows:

$$\rho^L_{u,t+1} \sim \beta_0 + \beta_L \delta_{u,t} + \vec{\beta}_X X_u + r_u + w_t + \rho^L_{u,t} \tag{5}$$

Here, $\rho^L_{u,t+1}$ represents the rate of cross-partisan posts containing style L (e.g., incivility, topics, linguistic markers) made by legislator u in week $t+1$, while $\rho^L_{u,t}$ denotes the corresponding rate in week t. The term $\rho^L_{u,t}$ controls for autoregressive dynamics, ensuring that baseline tendencies in style usage are accounted for when estimating the effect of engagement. The remaining variables are the same as RQ2.

Posting styles include both linguistic and affective features as described in section 4.1, except topics are represented as binary indicators denoting whether a given issue is present in a post, allowing us to assess the overall use of topical content in CPIs. The coefficient β_L measures the effect of engagement on the adoption of posting style L in subsequent cross-party communication. A positive coefficient indicates that legislators use of style L is positively associated with engagement of prior CPIs. We perform p-value correction (Bonferroni) for each interaction type and party, to account for multiple comparisons.

To satisfy assumptions of linear regression, all continuous variables are suitably transformed to be close to normal distributions and standardized for equations 2, 3, 4 and 5. Therefore, the coefficients should be interpreted as changes in standard deviation (SD) units.

5 Results

This section presents results for each RQ, reporting standardized effect sizes in the main text. Coefficients α and β reflect changes in standard deviations. Non-standardized effect sizes (Figs. 15–17) and qualitative examples (Table 5) in the Appendix provide additional percentage-based and contextual interpretation.

5.1 RQ1a: What is the Observed Effect of Cross-cutting Interactions on Engagement?

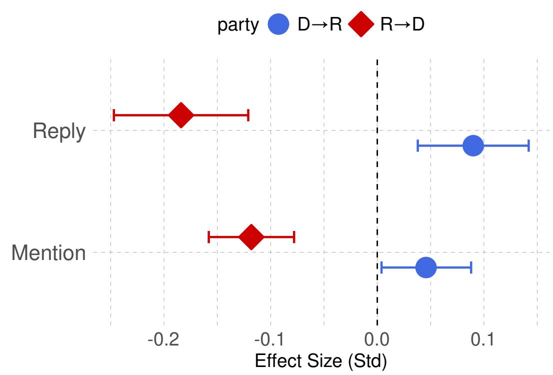

Figure 3: **Effect of CPIs on engagement.** Engagement of legislators varies significantly in cross-talks, with *asymmetric* effects depending on party affiliation and the nature of interactions. Effect size sign: (+) Higher (outperforming) engagement in CPI; (-) Lower (underperforming) engagement in CPI. Significant effects are denoted by filled markers ($p < 0.05$).

Figure 3 shows the effect sizes and 95% CI for the effect of cross-cutting interactions on engagement. When Republicans reply to (α_1 = -0.18, $p < 0.001$) or mention (α_1 = -0.12, $p < 0.001$) Democrats, their posts receive lower engagement compared to intra-party interactions. In contrast, Democrats gain higher engagement when replying to (α_1 = 0.09, $p < 0.001$) Republicans (corresponding to 8% engagement gains as shown in Fig. 15) or mentioning (α_1 = 0.05, $p < 0.05$) them. This may indicate that Republican audiences are less receptive to cross-party engagements through replies or mentions, potentially leading to reduced engagement. Conversely, the positive effects for Democrats suggest their audiences may favor more intentional and direct cross-party interactions.

Main Takeaways. Cross-cutting communication has an effect on the engagement of both Republican and Democrat legislators. Republican party audiences are less receptive to deliberative cross-party exchanges, while Democrat party audiences appear more interested in conversational engagements through mentions and replies, highlighting the asymmetries in political party audiences. These differences could also arise from the nature of these communications which we further explore in RQ1b.

5.2 RQ1b: What are the Observed Effects of Posting Styles on Engagement of CPIs?

Figure 4 presents how different posting styles influence the engagement of CPIs. Details on significance levels appear in Appendix Table 4, with illustrative examples of posts provided in Appendix Table 5.

Reply. In replies, engagement is strongly conditioned by both topic and tone. Democrats lose engagement when replying to Republicans about the Capitol riots ($\alpha_L = -0.89$). CPIs on such contentious topics, as shown in Appendix Table 5 (11, 14), often take the form of hostile rebuttals rather than substantive discussion, which likely discourages engagement. By contrast, Republicans gain engagement when their replies include incivility ($\alpha_L = 0.18$) or negative emotions such as anger ($\alpha_L = 0.09$) and anxiety ($\alpha_L = 0.14$). Examples like 15, 17 illustrate how emotionally charged replies can resonate with Republican audiences, even as they risk entrenching antagonistic rhetoric.

Mention. In mentions, Republicans lose engagement when raising the topic of abortion ($\alpha_L = -0.23$), whereas Democrats gain attention when mentioning Republicans on rights ($\alpha_L = 0.27$), gun control ($\alpha_L = 0.27$), or the Capitol riots ($\alpha_L = 0.25$). In Appendix Table 5, examples 24 and 26 illustrate how issue-centered mentions are rewarded, contrasting with combative replies. Both parties see engagement gains of about 13% when mentioning opponents in the context of COVID-19 (Appendix Fig. 16), suggesting bipartisan conversations on issues of national urgency resonate more broadly. Importantly, incivility in mentions reduces engagement for both Democrats (–8.3%) and Republicans (–21%), showing that audiences do not reward overt hostility in this interaction type. At the same time, the use of negative emotions modestly increases engagement: anger boosts engagement for both Democrats ($\alpha_L = 0.07$) and Republicans ($\alpha_L = 0.12$), while anxiety provides a boost for Democrats ($\alpha_L = 0.05$).

Other rhetorical markers also matter: URLs increase engagement, particularly for Democrats ($\alpha_L = 0.25$ vs. 0.06 for Republicans), while subjective language boosts engagement for both parties ($\alpha_L = 0.30$ for Democrats; 0.32 for Republicans). Hedging similarly lowers Republican engagement ($\alpha_L = -0.11$), suggesting their audiences respond unfavorably to uncertainty. Generalizations split audiences, enhancing Democratic mentions ($\alpha_L = 0.31$) but reducing Republican ones ($\alpha_L = -0.28$). Examples 29 and 30 highlight these contrasting uses of generalization language.

Main Takeaways. Posting styles are strongly associated with engagement of CPIs, with clear partisan asymmetries. In replies, Republican audiences reward antagonistic and emotional cues, while Democrats face engagement

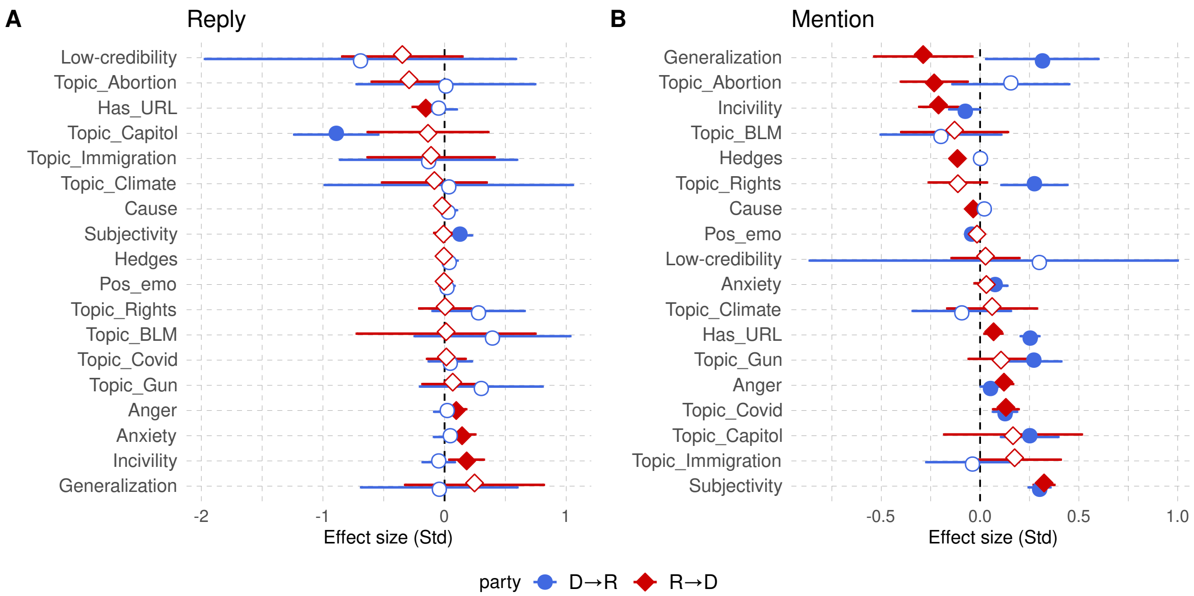

Figure 4: **Effect of posting styles on engagement of CPIs.** Posting styles are associated with the engagement garnered by CPIs and the associations differ across party and interaction types. The covariates are sorted based on effect sizes for Republicans.

losses when engaging on contentious issues like Capitol riots. In mentions, Democratic audiences engage more with issue-centered discussions, URLs, and generalized framing, whereas Republicans are penalized for incivility, uncertainty, and generalization language. Both parties benefit when addressing nationally salient issues like COVID-19. Together, these engagement patterns provide the foundation for understanding how audience responses may shape the frequency and style of elites' future CPI.

5.3 RQ2: What is the Observed Impact of Engagement on Future CPIs?

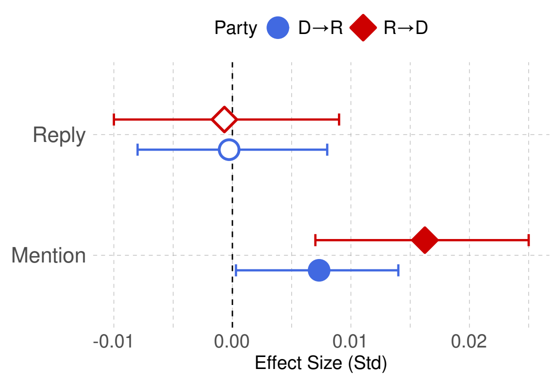

Figure 5: **Relationship between engagement and future CPIs.** Engagement with cross-partisan posts through mentions is strongly linked to increased future CPIs.

In RQ3, we examine how engagement from cross-cutting interactions is associated with the frequency of future cross-party communication by legislators. Figure 5 shows the effect sizes by type of interaction and party. For mentions, both Republicans (β_1 = 0.016, $p < 0.001$) (corresponds to 2.1% increase as shown in Fig. 17) and Democrats (β_1 = 0.007, $p < 0.05$) (1.3% rise) increase their cross-party mentions when these posts garner higher engagement. This suggests that legislators might recognize that engaging with the other side can amplify their reach and impact and hence tend to engage more in such cross-cutting discussions, more so in the case of Republicans. Though, in general, cross-party mentioning reduces the engagement of Republicans (RQ1), they tend to increase such interactions in the future when these garner higher attention, which highlights the subtleties of political communication online.

We do not observe significant effects for cross-cutting replies. This suggests a feedback loop in which political elites tailor their content based on audience preferences, but this effect is limited to mentions. One explanation could be that replies receive less engagement than mentions, making the small engagement gains from replies insufficient to motivate political elites to prioritize such interactions in future conversations.

Main Takeaways. Our findings point to a engagement-driven feedback loop in which audience reactions condition elite communication strategies. Legislators are more likely to sustain cross-party mentions after these interactions attract attention, suggesting that engagement acts as a reinforcement signal encouraging future bipartisan engagement. At the same time, the absence of effects for replies highlights that not all forms of dialogue generate the same incentives, underscoring the selective and complex ways in which online attention shapes elite cross-party behavior. Next, we examine whether engagement shapes not only the frequency of

CPIs but also the rhetorical and stylistic choices elites adopt in future CPIs.

5.4 RQ3: What is the Observed Impact of Engagement on the Style of Future CPIs?

Figure 6 presents how engagement gained from cross-party interactions is associated with the posting styles elites adopt in subsequent CPIs. We find systematic evidence that legislators adjust their linguistic and emotional strategies depending on whether CPIs previously attracted audience attention, consistent with the interpretation that engagement is correlated with stylistic differences in future cross-partisan communication.

Replies. In replies, parties exhibit shifts toward styles characterized by greater subjectivity and hedging following CPIs that receive higher engagement. For Democrats, engagement is associated with increased use of hedges ($\beta_L = 0.009, p < 0.05$) and subjective language ($\beta_L = 0.012, p < 0.001$), which are linguistic markers commonly used to soften claims, express personal viewpoints, and signal epistemic uncertainty. Republicans show a similar pattern, with greater use of subjective language ($\beta_L = 0.009, p < 0.05$) alongside more frequent use of causality terms ($\beta_L = 0.009, p < 0.05$). Taken together, these patterns indicate that higher engagement is associated with replies that rely more on subjective framing and explanatory language, rather than categorical or declarative statements.

Mentions. Mentions are associated with broader stylistic differences. Democrats are more likely to adopt subjective language ($\beta_L = 0.007, p < 0.05$) and express positive emotion ($\beta_L = 0.007, p < 0.05$) when prior mentions gain engagement, corresponding to more affectively warm and opinion-driven dialogue. Republicans show stronger associations: higher engagement is associated with reduced use of URLs ($\beta_L = -0.015, p < 0.001$), alongside increased use of hedges ($\beta_L = 0.011, p < 0.01$), subjective language ($\beta_L = 0.013, p < 0.01$), positive emotion ($\beta_L = 0.014, p < 0.001$), causality terms ($\beta_L = 0.01, p < 0.05$), and topical references ($\beta_L = 0.011, p < 0.01$). These associations suggest that highly engaged mentions tend to co-occur with cross-party communication that is more conversational, affective, and narrative-driven, relying less on external links.

Main Takeaways. We observe a pattern: Democrats exhibit more subjective and emotionally positive CPI, particularly in mentions and replies, while Republicans display broader stylistic differences that include causal reasoning and issue-centered framing. In deliberative forms of CPI—replies and mentions—higher engagement is systematically associated with styles linked to more meaningful discourse: hedging and causal explanations that provide reasoning, subjective framing that signals personal engagement, and positive-emotional language.

We do not find evidence that engagement is associated with increased use of negative emotions such as anger, anxiety, or toxicity in future CPIs—despite these styles garnering high engagement for certain CPIs (RQ1b). This absence is noteworthy: while prior work raises concerns that engagement metrics may incentivize divisive or hostile rhetoric, our results indicate that, in the context of CPIs, higher engagement tends to co-occur with more constructive and deliberative communication strategies. These patterns suggest that audience engagement is correlated with cross-party exchanges that are more deliberative, substantive.

5.5 Robustness Checks

We perform several checks to assess whether our results are robust to alternative specifications and design choices. For RQ1, we try different matching strategies, including propensity-score and nearest-neighbor matching with varying covariate sets, and find consistent results across specifications. For RQ2 and RQ3, we reestimate models using monthly and biweekly instead of weekly aggregates, and the association between engagement and both the frequency and style of subsequent CPIs remains unchanged. We also test alternative mixed-effects model specifications, varying autoregressive terms, user-level random effects, and time fixed effects. These checks yield results that closely mirror the main analyses, with no changes in the direction of effects and only minor variation in magnitude (within 2-5% of reported results). These checks suggest that our findings are not artifacts of modeling or sampling choices but reflect robust patterns in how engagement dynamics shape elite CPI.

6 Discussion

Our findings show that audience reactions are not merely passive reflections of elite communication, but actively shape how elites engage in cross-partisan discourse. While prior work often frames social media as amplifying polarization, our results point to a more nuanced dynamic in which audience engagement can encourage continued cross-party interaction and shifts in rhetorical strategy. Highly visible cross-partisan exchanges are associated with increased use of causal reasoning, subjective framing, and positive-emotion language in subsequent interactions. These patterns suggest that engagement incentives may, under certain conditions, support more constructive forms of discourse even within politically contentious environments.

These results contribute to a broader understanding of online political communication as a socio-technical feedback system in which platform incentives and behavioral responses jointly shape elite expression. Rather than treating elite communication as independent of audience reactions, our findings show that elites adapt their communication strategies in response to engagement signals. This perspective highlights how visibility metrics and interaction dynamics influence not only what content spreads, but also how political actors talk across partisan boundaries.

More broadly, this work helps reconcile seemingly conflicting findings in prior literature by showing that social media environments can simultaneously create pressures toward polarization while also sustaining opportunities for cross-party interaction. Our results suggest that engagement does not uniformly reward divisive communication, but may

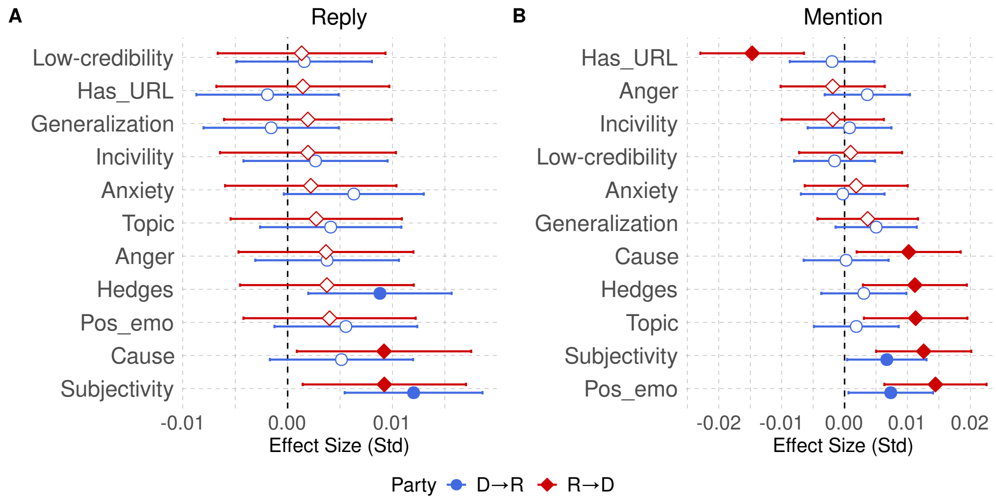

Figure 6: **Association between engagement and future cross-talk style.** Prior CPI engagement is associated with differences in the rhetorical and stylistic choices employed by legislators in their subsequent cross-cutting interactions. Both Republicans and Democrats exhibit shifts toward more constructive CPI styles, as indicated by greater use of subjective, causal, and positive emotional language.

also reinforce rhetorical strategies associated with explanation, interpretation, and affective framing. Taken together, these findings underscore the importance of studying elite behavior as dynamically embedded within interactive systems shaped by audience attention, platform design, and strategic adaptation.

Limitations and Future Work. Our analysis provides new evidence of feedback loops in elite cross-party communication, but certain limitations remain. We focus on Twitter during 2020–21, a period shaped by specific platform affordances and events; patterns of engagement and reinforcement may differ elsewhere. While we measure engagement relative to legislators' baselines, the lack of granular audience identifiers limits our ability to separate co-partisan from cross-partisan drivers. Importantly, our analyses are observational and describe associations, and we do not claim any causal effects. As with most large-scale observational work, our findings should be interpreted as descriptive associations, with multiple robustness checks supporting the stability of these patterns across alternative specifications.

Future work should extend this analysis across platforms, election cycles, and political systems to test the generalizability of feedback effects. Experimental and mixed-method approaches could further clarify how audiences interpret and reward rhetorical strategies, deepening understanding of when engagement fosters bridging vs. division.

Acknowledgements

The authors would like to acknowledge support from AFOSR, ONR, Minerva, NSF #2318461, and Pitt Cyber Institute's PCAG awards. The research was partly supported by Pitt's CRCD resources. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.

References

An, J.; et al. 2019. Political discussions in homogeneous and cross-cutting communication spaces. In *ICWSM*, volume 13.

Bail, C. A.; et al. 2018. Exposure to opposing views on social media can increase political polarization. *PNAS*.

Barber, B. R. 2003. *Strong democracy: Participatory politics for a new age*. Univ of California Press.

Bishop, S. 2019. Managing visibility on YouTube through algorithmic gossip. *New media & society*, 21(11-12).

Biswas, A.; et al. 2025. Political elites in the attention economy: Visibility over civility and credibility? In *ICWSM*, volume 19, 241–258.

Burrell, M. 2020. *Why Would I Want to Talk to Them? An Exploration of Perceptions of Talking across Political Divides*. Nova Southeastern University.

Caplan, R.; and Boyd, D. 2016. Who controls the public sphere in an era of algorithms. *Mediation, Automation, Power*, 1–19.

Carrubba, C. J. 2001. The electoral connection in European Union politics. *The Journal of politics*, 63(1): 141–158.

Cinelli, M.; et al. 2021. The echo chamber effect on social media. *PNAS*, 118(9): e2023301118.

De Castella, K.; and McGarty, C. 2011. Two leaders, two wars: A psychological analysis of fear and anger content in political rhetoric about terrorism. *ASAP*, 11(1): 180–200.

Demszky, D.; et al. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

Downs, A. 1957. An economic theory of democracy. *Harper and Row*, 28.

Flores-Saviaga, C.; et al. 2018. Mobilizing the trump train: Understanding collective action in a political trolling community. In *ICWSM*, volume 12.

Frimer, J. A.; et al. 2023. Incivility is rising among American politicians on Twitter. *SPPS*, 14(2): 259–269.

Gabel, M.; and Scheve, K. 2007. Estimating the effect of elite communications on public opinion using instrumental variables. *AJPS*, 51(4): 1013–1028.

Garimella, K.; et al. 2018. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 world wide web conference*.

Garimella, V. R. K.; and Weber, I. 2017. A long-term analysis of polarization on Twitter. In *ICWSM*, volume 11.

Goovaerts, I.; et al. 2020. Uncivil communication and simplistic argumentation: Decreasing political trust, increasing persuasive power? *Political Communication*, 37(6).

Graham, J.; et al. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5): 1029.

Grossmann, M.; and Hopkins, D. A. 2016. *Asymmetric politics: Ideological Republicans and group interest Democrats*. Oxford University Press.

Habermas, J. 1989. The structural transformation of the public sphere, trans. Thomas Burger. *Cambridge: MIT Press*.

Hanley, H. W.; and Durumeric, Z. 2023. Sub-Standards and Mal-Practices: Misinformation's Role in Insular, Polarized, and Toxic Interactions. *arXiv preprint arXiv:2301.11486*.

Hasan, R.; et al. 2022. The Impact of Viral Posts on Visibility and Behavior of Professionals: A Longitudinal Study of Scientists on Twitter. In *ICWSM*, volume 16, 323–334.

Heatherly, K. A.; et al. 2017. Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites. *New Media & Society*, 19(8): 1271–1289.

Hua, Y.; et al. 2020. Towards measuring adversarial twitter interactions against candidates in the us midterm elections. In *ICWSM*, volume 14, 272–282.

Islam, J.; et al. 2020. A lexicon-based approach for detecting hedges in informal text. In *LREC*, 3109–3113.

Jacoby, J. 2002. Stimulus-organism-response reconsidered: an evolutionary step in modeling (consumer) behavior. *Journal of consumer psychology*, 12(1): 51–57.

Jost, J. T.; et al. 2007. Are needs to manage uncertainty and threat associated with political conservatism or ideological extremity? *Personality and social psychology bulletin*.

Jun, N. 2014. Political diversity and participation: a systematic review of the measurement and relationship. *Asian Journal for Public Opinion Research*, 1(2): 103–127.

Kim, J. W.; et al. 2021. The distorting prism of social media: How self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication*, 71(6).

Kim, Y.; and Chen, H.-T. 2016. Social media and online political participation: The mediating role of exposure to cross-cutting and like-minded perspectives. *Telematics and Informatics*, 33(2): 320–330.

Lasser, J.; et al. 2022. Social media sharing of low-quality news sources by political elites. *PNAS nexus*, 1(4): pgac186.

Liu, Y.; et al. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Lu, Y.; et al. 2016. Cross-cutting exposure on social networking sites: The effects of SNS discussion disagreement on political participation. *Computers in Human Behavior*.

Marchal, N. 2022. "Be nice or leave me alone": An intergroup perspective on affective polarization in online political discussions. *Communication Research*, 49(3): 376–398.

Min, S. J.; and Wohn, D. Y. 2018. All the news that you don't like: Cross-cutting exposure and political participation in the age of social media. *Computers in Human Behavior*.

Morales, D. F.; et al. 2021. No echo in the chambers of political interactions on Reddit. *Scientific reports*, 11(1): 2818.

Mutz, D. C. 2001. Facilitating communication across lines of political difference: The role of mass media. *American political science review*, 95(1): 97–114.

Mutz, D. C. 2002. The consequences of cross-cutting networks for political participation. *AJPS*, 838–855.

Mutz, D. C. 2007. Effects of "in-your-face" television discourse on perceptions of a legitimate opposition. *APSR*.

Nyhan, B.; and Reifler, J. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior*.

Pettigrew, T. F.; and Tropp, L. R. 2006. A meta-analytic test of intergroup contact theory. *JPSP*, 90(5): 751.

Phillips, W. 2015. *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. Mit Press.

Ryan, T. J. 2012. What makes us click? Demonstrating incentives for angry discourse with digital-age field experiments. *The Journal of Politics*, 74(4): 1138–1152.

Sahly, A.; et al. 2019. Social media for political campaigns: An examination of Trump's and Clinton's frame building and its effect on audience engagement. *Social Media+ Society*, 5(2).

Santoro, L. R.; et al. 2021. Exploring the direct and indirect effects of elite influence on public opinion. *Plos one*, 16(11).

Saveski, M.; et al. 2022. Engaging politically diverse audiences on social media. In *ICWSM*, volume 16.

Shor, B.; and McCarty, N. 2011. The ideological mapping of American legislatures. *APSR*, 105(3): 530–551.

Slater, M. D. 2007. Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication theory*, 17(3): 281–303.

Taber, C. S.; and Lodge, M. 2006. Motivated skepticism in the evaluation of political beliefs. *AJPS*, 50(3): 755–769.

Tai, Y. C.; et al. 2023. Official yet questionable: examining misinformation in US state legislators' tweets. *J. Inf. Technol. Politics*, 1–13.

Tufekci, Z. 2013. "Not this one" social movements, the attention economy, and microcelebrity networked activism. *American behavioral scientist*, 57(7): 848–870.

Wiebe, J.; et al. 2000. Learning subjective adjectives from corpora. *Aaai/iaai*, 20(0): 0.

Wilson, T.; et al. 2005. OpinionFinder: A system for subjectivity analysis. In *Proceedings of HLT/EMNLP 2005 Interactive Demonstrations*, 34–35.

Wu, S.; and Resnick, P. 2021. Cross-Partisan Discussions on YouTube: Conservatives Talk to Liberals but Liberals Don't Talk to Conservatives. In *ICWSM*, volume 15, 808–819.

Yu, X.; et al. 2024. Partisanship on social media: In-party love among American politicians, greater engagement with out-party hate among ordinary users. *Political Behavior*.

Zaller, J. 1992. *The nature and origins of mass opinion*. Cambridge university press.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes. See Section 5.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes. See Section 4.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes. See Section 3.
   (e) Did you describe the limitations of your work? Yes. See Section 6.
   (f) Did you discuss any potential negative societal impacts of your work? Yes. See Section 6, 7.
   (g) Did you discuss any potential misuse of your work? Yes. We discuss potential mis-intepretation of our conclusions in Section 6.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes. We will release the dataset in the future.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes. See Section 2.
   (b) Have you provided justifications for all theoretical results? Yes. See Sections 6 and 7.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes. See Section 2, 4.
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes. Throughout the study we have made considerations to address for potential confounding and conducted robustness checks.
   (e) Did you address potential biases or limitations in your theoretical framework? Yes. See Sections 6, 7.
   (f) Have you related your theoretical results to the existing literature in social science? Yes. See Sections 2, 4, 6.
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes. We discuss the implications in Section 6.
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? N/A.
   (b) Did you include complete proofs of all theoretical results? N/A.
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? N/A.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? N/A.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? N/A.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? N/A.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? N/A.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? N/A.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes. See Section 3.
   (b) Did you mention the license of the assets? N/A.
   (c) Did you include any new assets in the supplemental material or as a URL? No.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes. See Section 7.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes. We conform to the social media policies. See Section 7.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? No. We need to conform to social media platform policies sharing our curated data. That means, we can only share the public post IDs, without the data content itself.
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? We will do our best releasing the data without breaching the social media platforms' policies.
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? N/A.
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? N/A.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? N/A.
   (d) Did you discuss how data is stored, shared, and deidentified? N/A.

Appendix

Note: For all manual labeling tasks reported in this paper, annotations were conducted independently by two annotators, and final labels were determined through discussion to resolve any disagreements.

Ethical Considerations. We collected data from Twitter using Twitter's Official API v2.0 before rate limitations were imposed. The data are posted by public figures and available for viewing without any restrictions. The study is performed at an aggregate level and we do not analyze or report any results at the individual level. Examples of actual tweets are only shared in Table 5 where we removed all personally identifiable information. Moreover, the results and implications presented in this work should not be interpreted by political elites as a means to boost their engagements on social media.

Engagement metric for retweets. Engagements on retweets are influenced by the engagement of the original post and its author. The engagement metric for retweets are limited to retweet counts, which inherently reflect the popularity of the original post. This limitation arises because the Twitter API attributes retweet engagement metrics (i.e., retweet counts) to the original post rather than to the retweeted post itself. Given these constraints, it is not feasible to fully disentangle the independent engagement dynamics of retweets versus the original posts. However, as legislators often have large follower bases, it is likely that their retweets amplify the engagement of the original posts, making the study of cross-party retweets an important part of our study.

The use of the overperforming score metric (Section 3.3) mitigates some of the concerns regarding the inability to separate retweet and original post engagements by ensuring that engagements (even across retweets) are analyzed relative to a legislator's baseline engagement levels. We analyze the correlation between the mean overperforming scores of a legislator's original posts and their retweeted posts, finding a significant positive Pearson correlation (0.38, $p < 0.05$). While modest, this correlation demonstrates that engagement dynamics for retweeted posts align with those of a legislator's original posts, relative to their baseline engagement levels. This alignment suggests that retweeted posts are not outliers or fundamentally different in terms of engagement patterns. Therefore, the overperforming score metric remains a valid tool for capturing engagement dynamics across post types. Although the metric in terms of retweets does not fully isolate the direct influence of the original post's engagement, the significant result asserts minimal biases from external factors, providing a reliable basis for analysis.

Cross-cutting interactions. We manually label 300 followers[20] for their political leaning (i.e., Democrat vs. Republican) using stratified sampling that intentionally oversamples users with more extreme attention-leaning scores (i.e., above 0.5 and below -0.5). Our goal is not to infer partisan leaning for all users, but to calibrate reliable thresholds for identifying users with clear Republican or Democratic leanings. Users near the center of the distribution frequently follow legislators from both parties and often lack a consistent partisan stance in their own content, making them inherently ambiguous and unsuitable for threshold calibration. During the manual labeling process, we analyze (at least five) recent posts per account to determine the political leaning of the users. An account is labeled as Republican or Democrat if the majority of its posts express a clear stance supporting or opposing a specific party, policy, or ideology. If the posts lack a consistent political stance or focus on non-political topics, the account is classified as non-partisan or apolitical. This likely ensures that the classifications are grounded in the user's behavior, reducing the likelihood of misclassifications due to ambiguous following patterns.

Two annotators labeled the samples with a Cohen Kappa score of 0.78 (substantial agreement). The final labels were decided by mutual discussion. Based on the ROC curves (Figure 7), we set the threshold (by maximizing TPR and minimizing FPR simultaneously) for identifying a follower as Democrat (AUC = 0.86) at 0.71 (i.e., scores between 0.71 and 1) and as Republican (AUC = 0.81) at -0.69. For Democrat leaning accounts, the precision, recall and F1 scores are 0.79, 0.85 and 0.82 respectively at the chosen cutoff. For Republican leaning accounts, the precision, recall and F1 scores are 0.81, 0.78 and 0.80 respectively.

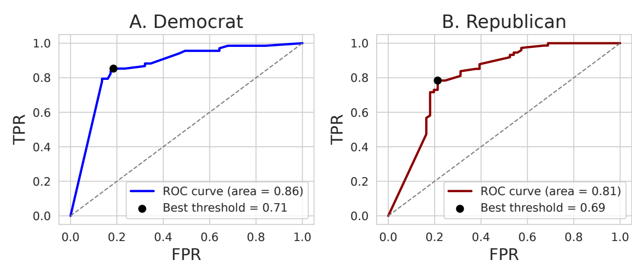

Figure 7: ROC curves for (A) Democrats and (B) Republicans showing optimal thresholds to determine user leanings.

Distribution of cross-party posts and their engagement. Figure 8 shows the distribution of cross-cutting and intra-partisan posts and their engagement.

Central (C) vs. local (L). We identify the location of the legislator's followers using the user location field returned by the Twitter API. This field is user-generated containing noisy data. We use regular expressions to identify user's state if they are from the US else their country. Around 46% followers are mapped to the states. The locality of a legislator's followers is measured as the proportion (calculated out of the followers having a location) of followers belonging to the same state as the legislator. The distribution of local and central legislators are similar for Republicans and Democrats, suggesting that no (or minimal) biases are introduced for party using our categorization method.

[20]The median number of followers per legislator is 489 (with maximum following of 24,598). We report median since the distribution of followers is skewed.

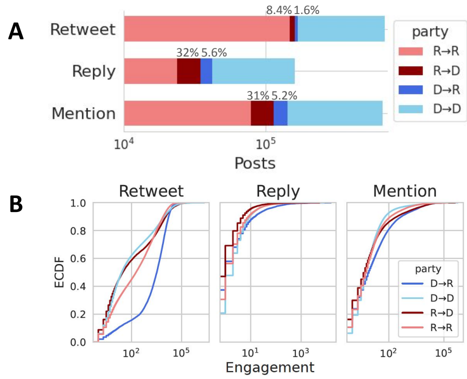

Figure 8: **Distribution of (A) cross-party posts and their (B) engagement.** The rate of cross-talk is higher for Republicans compared to Democrats across all types of interactions. However, typical engagements on cross-party posts tend to be lower than intra-party ones for Republicans.

Measuring Engagement. We borrow the engagement metric and the suggested parameters $b_0 = 10$ and $m = 14$, from Biswas et al. (2025), who demonstrated these parameters to be effective in capturing meaningful engagement variations. The b_0 is based on the daily mean engagement received by legislators on Twitter. It serves as a threshold to filter out noisy engagement posts by ensuring that only posts with substantial engagement levels are considered, thereby reducing the impact of low-engagement anomalies on the engagement metric. The ideal window m is estimated based on legislators' daily posting rates on these platforms. Choosing $m = 14$, ensures a reliable estimate of expected engagement b_u by incorporating a sufficient number of posts per legislator while accounting for temporal variations in posting behavior and audience activity.

These design choices are likely to make the engagement metric robust and meaningful across varying levels of activity among legislators. While the confidence in the median engagement b_u may be lower for legislators with fewer posts, this is likely mitigated by the choice of $m = 14$, which is derived from the typical posting rates of legislators. This ensures that, on average, we have a sufficient sample size per legislator for a reliable calculation of b_u. Additionally, our analysis does not rely on the cumulative engagement of all posts but instead focuses on the median engagement for each legislator during the preceding window. This approach inherently controls for variations in posting frequency, as it standardizes engagement levels across legislators regardless of their activity rates, reducing potential biases introduced by differences in the volume of posts.

Precision of Linguistic Markers. We manually label 40 random samples per linguistic marker (i.e., 280 posts overall) to evaluate the quality of our lexicon-based features. Annotators showed high agreement (Cohen's κ = 0.82), with disagreements resolved through discussion to produce final reference labels. Table 2 reports the precision of the lexicon-based indicators with respect to these human-validated labels, both overall and by party. The overall precision ranges between 0.625-0.925, indicating that when a linguistic marker is identified by the lexicon-based method, it typically corresponds to the intended construct as judged by human annotators. We focus on precision rather than recall because our goal is to use conservative, high-confidence indicators of linguistic traits at scale, rather than to exhaustively capture all possible instances of a given style. The precision of certain markers, such as `cause` and `pos_emo`, differs across parties, which may reflect differences in communication styles.

	Anger	Anxiety	Cause	Generalize	Hedges	Pos_emo	Subj
Overall (Manual)	0.625	0.775	0.725	0.925	0.675	0.775	0.825
Dem (Manual)	0.643	0.806	0.667	0.923	0.620	0.710	0.786
Rep (Manual)	0.583	0.667	0.900	0.928	0.818	1.000	0.917
LLaMA-3-70B	0.731	0.782	0.820	0.846	0.680	0.825	0.715

Table 2: Precision of Linguistic Markers

LLM-based validation of posting style annotations. To further assess the reliability of our lexicon-based posting style indicators, we conducted an additional validation using a large language model (LLM). For these 280 posts, we prompted LLaMA-3-70B (Instruct) to judge whether each post exhibited specific linguistic characteristics (e.g., subjectivity, causality, hedging, incivility) using the same definitions as our lexicon-based measures. We then compared the LLM's judgments with the binary labels produced by the lexicon-based approach. We observe good agreement between the two methods across most posting styles, as reported in Table 2. Disagreements primarily arise in cases where pragmatic interpretation is required. For instance, the LLM sometimes infers sarcasm, implied blame, or emotional intent in posts that contain no explicit linguistic markers, leading to false positives. In contrast, the lexicon-based approach is more conservative, labeling a post only when explicit lexical cues are present. Given these trade-offs, we retain lexicon-based indicators as our primary measures for scalability and interpretability.

Topics. Our topic selection approach is guided by the study's focus on understanding how salient topics in cross-party discussions affect the engagement of legislators, rather than on an exhaustive exploration of all possible topics in cross-party discussions. The topics selected (BLM, COVID-19, rights, immigration, gun control, climate, abortion, and the Capitol riots) represent highly salient and contentious issues in contemporary U.S. politics. While not exhaustive, this selection ensures coverage of areas that are critical for analyzing US political discussions. By focusing on these topics, we aim to capture the dynamics of cross-party interactions around some of the most significant issues of public and political interest.

We use a keyword-based approach for initial topic detection because it offers interpretability and ensures high precision. This method is particularly effective given the brevity of tweets, which often limits the performance of unsupervised techniques like BERTopic. While we initially experiment with BERTopic to identify topics in an unsupervised manner, the results are noisy (and especially ineffective due to our focus on a specific set of topics), likely due to the short and informal nature of tweets. Consequently, we adopt a more structured and precise methodology to identify topical tweets.

An initial set of keywords (shown in Table 3) specific to each topic is used to identify topical tweets. While this set may not be comprehensive, it is designed to maximize precision over recall, thereby reducing false positives. This ensures that the posts identified for a topic are highly relevant, even if some relevant posts are excluded. Tweets identified for each topic are clustered based on their RoBERTa embeddings (using Nearest Neighbor) to capture semantic similarity. This step gives clusters of posts under each topic which is used to identify relevant clusters for each topic. We manually label 25 posts per cluster to identify and exclude noisy clusters under each topic. The clusters containing posts irrelevant to the topic are filtered out. Finally, we assign all posts belonging to relevant clusters to the seed topic. An additional 50 posts per topic (after filtering out noisy clusters) are manually labeled to ensure the relevance of classified posts. The precision of our topic labeling ranges between 84-100%, demonstrating the reliability of the approach. Based on our methodology, we are able to assign posts to multiple topics despite the short text, enabling nuanced analysis of complex discussions. Table 3 shows the distribution of topics for cross-cutting and co-partisan posts[21].

Low-credibility. Among the low-credibility URLs[22], 38.1% are news-related. Using URLs to detect low-credibility content is a common practice in the literature (Lasser et al. 2022), particularly for short texts like tweets, as it provides a straightforward method to evaluate the reliability of shared information. While other methods, such as detecting fact-checked claims using LLMs, are promising, they present challenges in our context. Our dataset spans an older time period, during which fact-checking datasets may be sparse. Furthermore, LLM-based approaches risk hallucination and lack robust interpretability, making validation without proper ground truth data more difficult. This underscores the practicality of our current approach, although future work could explore integrating newer methodologies as tools and datasets improve.

It is important to note that our work is not focused on quantifying the prevalence of low-credibility content overall. Instead, it explores the role of low-credibility content as a specific posting style in the engagement dynamics of cross-cutting interactions. Thus, we are only interested in how the use of such content may influence interactions between opposing partisan groups, rather than the amount or nature of low-credibility content legislators share.

[21] Previously, the topic 'race' was also included but the clusters for this topic were noisy, hence we removed it for further analysis

[22] Around 11.1% of posts include URLs. We are unable to determine how many of these URLs are news-related, as recovering full URLs for all posts is non-trivial. This limitation arises because URLs in Twitter posts are often truncated, requiring additional data collection or third-party services to expand and categorize them reliably.

Mann-Whitney U test. Figure 9) shows the Mann-Whitney effect sizes for differences in cross-cutting vs. intra-party posts (Figure 10 shows the confidence intervals).

Matching. We employ 1:1 matching using K-nearest neighbor algorithm (with Euclidean distance) for RQ1a and 1b. The distance cutoff is chosen such that the covariate overlap between treated and control groups are maximized. For cross-cutting posts by Republicans we find matches for 73.2% retweet, 74.7% reply, and 75.3% mentions. For Democrats, we find matches for 78.2% retweet, 76.2% reply, and 78.8% mentions. All the covariates are balanced after matching as shown in Fig. 11 (bootstrapped mean and 95% confidence intervals for standardized differences after matching are shown in Fig. 12).

Data Transformation for RQs. The variables were transformed to be close to normal distributions to satisfy the model assumptions. For RQ1a and 1b, the transformations are: y (log), audience partisanship (arcsinh), centrality (center+scale), #posts (log), #followers (Yeo-Johnson), and ideology (arcsinh). For RQ2 and 3, the transformations are: ρ (log), δ (arcsinh), ideology (sqrt), #followers (arcsinh), #posts (Yeo-Johnson), centrality (Yeo-Johnson), and audience partisanship (Yeo-Johnson).

Effect of cross-talk for Retweet. Figure 13 shows the effect of cross-talk on engagement (RQ1) for Retweet.

As shown in Figure 13(a), Republicans get higher engagement when they retweet Democrats (α_1 = 0.16, $p < 0.001$) i.e., a post tends to overperform by 16%, on average, if a Republican legislator retweets the post by a Democrat. However, Democrats get lower engagement on retweeting Republicans (α_1 = -0.12, $p < 0.05$). This suggests that Republican audiences respond more positively to amplifying cross-cutting content, leading to increased engagement when Republicans engage with Democratic posts. On the other hand, Democratic audiences tend to engage less with cross-cutting content, leading to decreased engagement when Democrats retweet Republican posts. This asymmetry reflects the differences in how the two groups interact with content that presents opposing viewpoints.

Figure 13(b) shows that topics like abortion ($\alpha_L = -0.84$) and gun control ($\alpha_L = -0.23$) reduce the engagement of Republicans when they retweet Democrats. In contrast, Democrats gain higher engagement when retweeting Republicans on issues such as gun control ($\alpha_L = 2.02$) and BLM ($\alpha_L = 0.93$). One possible explanation is that Democrats' retweets tend to focus more on specific issues (e.g., sample tweets ④ and ⑥ in Appendix Table 5)), while Republicans are more likely to retweet Democratic content sarcastically (e.g., ① and ③ in Appendix Table 5). Notably, the presence of anger is associated with higher engagement for Democrats (α_L = 0.29). These cross-party retweets tend to be stances expressed about certain issues (e.g., ⑩ in Appendix Table 5), suggesting that such posts resonate with Democratic audiences.

Feedback Loop Effect for Retweet. Figure 14 shows the effect of engagement on future cross-talk rate (RQ2) and

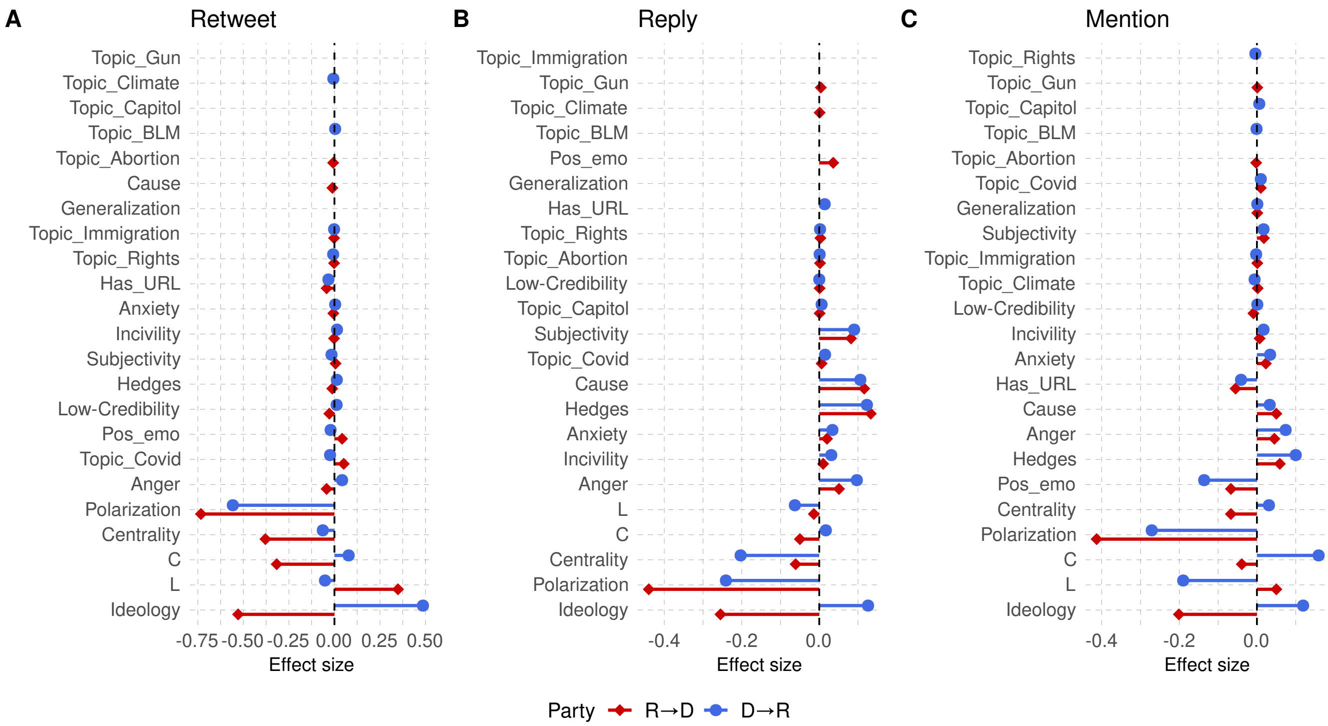

Figure 9: **Differences in Posting Styles and Individual Characteristics.** The posting styles and individual attributes of legislators vary between cross-party and intra-party posts, for both political parties and across both interaction types. The figure uses lollipop charts to show the differences between cross-party and intra-party (baseline) posts, measured by effect sizes through the Mann-Whitney U test. Only statistically significant effects ($p < 0.05$) are displayed. The covariates are sorted based on differences in effect sizes between Republicans and Democrats (low to high). Effect size sign: (+) More frequent in cross-party group; (-) More frequent in intra-party group.

stylistic choices (RQ3) for Retweet. Republican legislators are more likely to retweet Democrats by 2.2% (β_1 = 0.018, $p < 0.001$) in the future when cross-party retweets gain higher engagement as shown in Figure 14(a). However, no significant effects are seen for Democrat elites. This is aligned with the results of RQ1 which suggests that retweets are a preferred mode of cross-partisan interaction for Republican audiences, unlike Democrats. Therefore, the gain in engagement from cross-party retweets motivates Republican legislators to further engage in such interactions in the future.

Retweets show the strongest evidence of stylistic reinforcement among Republicans as shown in 14(b). For Democrats, visible CPIs lead to increased use of toxicity ($\beta_L = 0.0080, p < 0.05$) and causality terms ($\beta_L = 0.0074, p < 0.05$), suggesting that engagement pushes them toward sharper and more explanatory retweeting of opponents. Republicans exhibit a wide array of reinforced styles: engagement predicts higher use of hedges ($\beta_L = 0.0211, p < 0.001$), subjective language ($\beta_L = 0.0107, p < 0.05$), positive emotion ($\beta_L = 0.0238, p < 0.001$), anger ($\beta_L = 0.0113, p < 0.01$), causality terms ($\beta_L = 0.0260, p < 0.001$), and topics ($\beta_L = 0.0259, p < 0.001$). Importantly, engagement is negatively associated with misinformation content ($\beta_L = -0.0123, p < 0.01$), suggesting that audiences discourage Republican elites from retweeting misleading or unreliable material. Together, these results show that retweets act as a particularly strong site of stylistic feedback loops, with Republicans especially likely to double down on emotional and issue-focused styles when their retweets gain traction.

RQ1b: p-value and examples. Table 4 shows the regression coefficients and corresponding p-values for RQ1b. Table 5 shows examples of some of the posting styles associated with the engagement of cross-cutting posts.

Non-standardized Regression Results. Figures 15, 16, 17 and 18 show the non-standardized effect sizes for RQ1a, 1b, 2, and 3 respectively. The non-standardized effect sizes (multiplying the coefficients by 100) can be interpreted as percentage changes in the outcome variable.

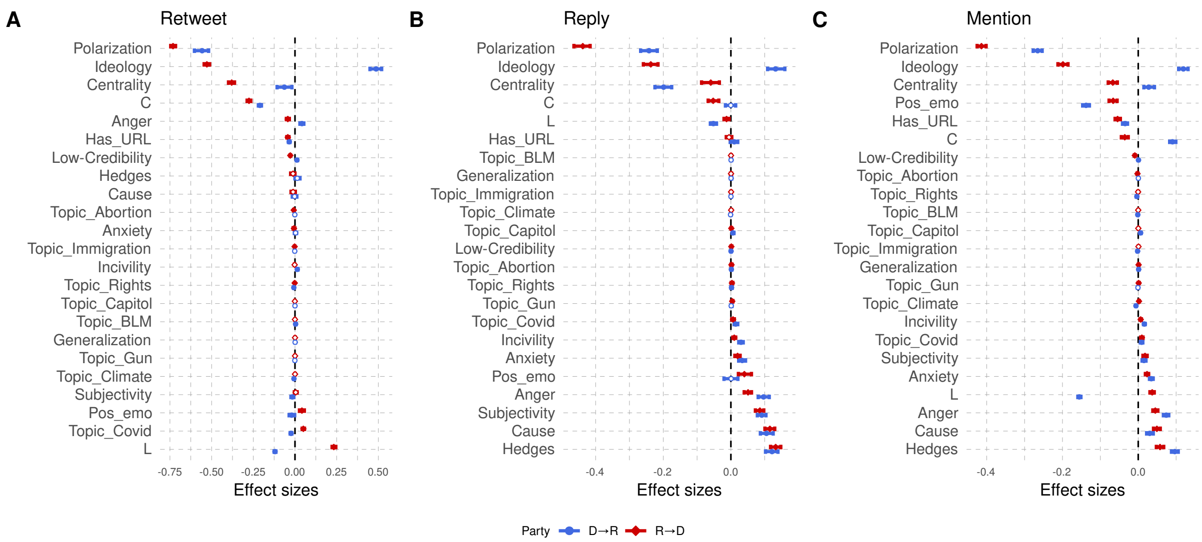

Figure 10: Confidence intervals of Mann-Whitney test (Figure 9). The covariates are sorted based on effect sizes for Republicans.

Topic	Initial Keyword(s)	cross-partisan R→D	cross-partisan D→R	co-partisan R→R	co-partisan D→D
BLM	george, floyd, blacklivesmatter	88 (0.19%)	123 (0.38%)	489 (0.22%)	3,825 (0.44%)
COVID-19	covid, covid-19, pandemic	2,457 (5.37%)	1,826 (5.59%)	6,885 (3.14%)	50,743 (5.79%)
rights	right(s)	287 (0.63%)	228 (0.70%)	1,128 (0.51%)	10,595 (1.21%)
immigration	immigration, immigrating, immigrant(s)	108 (0.24%)	78 (0.24%)	687 (0.31%)	3,950 (0.45%)
gun control	gun(s), shooting	313 (0.68%)	267 (0.82%)	1,139 (0.52%)	7,636 (0.87%)
climate	climate	141 (0.31%)	74 (0.23%)	303 (0.14%)	8,373 (0.96%)
abortion	abortion	1,733 (3.79%)	90 (0.28%)	1,733 (0.79%)	2,825 (0.32%)
Capitol riots	capitol, riot(s)	76 (0.17%)	277 (0.85%)	413 (0.19%)	2,187 (0.25%)

Table 3: Keywords used to identify topics and distribution of topics for cross-partisan and co-partisan posts. Percentages denote the share of posts in each interaction category containing the given topic.

	Retweet		Reply		Mention	
	R→D	D→R	R→D	D→R	R→D	D→R
Abortion	**-0.836*****	0.574	**-0.291**†	0.010	**-0.232****	0.155
BLM	-0.030	**0.935****	0.013	0.393	-0.129	-0.198
Capitol riots	0.077	-0.192	-0.136	**-0.891*****	0.166	**0.251*****
Climate	-0.109	0.591	-0.084	0.035	0.061	-0.092
Covid	0.016	0.063	0.014	0.047	**0.131*****	**0.127*****
Gun	**-0.229**†	**2.020*****	0.067	0.302	0.106	**0.272*****
Immigration	0.044	0.038	-0.111	-0.131	0.175	-0.038
Rights	-0.147	-0.384	0.004	0.279	-0.112	**0.275****
Has_URL	-0.005	0.111	**-0.154****	-0.049	**0.068****	**0.253*****
Incivility	**-0.170**†	0.117	**0.181***	-0.048	**-0.210*****	-0.075
Generalization	0.092	-0.306	0.246	-0.043	**-0.288***	**0.315***
Hedges	-0.007	0.061	-0.005	0.038	**-0.114*****	0.001
Subjectivity	**0.068****	-0.066	-0.009	**0.127***	**0.323*****	**0.301*****
Pos_emo	**-0.043*****	0.025	-0.005	0.019	-0.015	**-0.042****
Anxiety	0.007	0.103	**0.144***	0.047	0.032	**0.076***
Anger	-0.003	**0.290*****	**0.095***	0.023	**0.121*****	**0.052***
Cause	0.018	-0.070	-0.019	0.029	**-0.034***	0.021
Low-Credibility	0.067	0.154	-0.347	-0.691	0.027	0.299

† $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression coefficients for RQ1b

Interaction	Feature	Examples	No.
Retweet	Abortion	(-) RT @xx: Seriously, though: for most of the religious right the "but Judges..." argument boils down to abortion and LGBT rights.	①
		RT @xx: I asked @xx three times this afternoon if he thinks something like Texas's 6-week abortion ban should be	②
	Gun	(-) RT @xx: In Texas you have to show strict ID to vote but not to carry a gun	③
		(+) RT @xx: Mass shootings and more heartbreak for families that lose loved ones to gun violence is a crisis in America.	④
	BLM	RT @xx: President @xx has ordered the FBI to expedite its investigation into the death of George Floyd.	⑤
		(+) RT @xx: George Floyd appears to have been the victim of murder. A close review of the video can lead one to no other conclusion	⑥
	Incivility	(-) RT @xx: Who is this white dude and why the fuck did he just ask this sista to give him examples of racism?	⑦
		RT @xx: Donald Trump is the stupidest	⑧
	Anger	RT @xx: We have the best military and the best intelligence anywhere in the world. If Americans are threatened, we are prepared..	⑨
		(+) RT @xx: This is testament the importance of All Americans participating in the Democratic process. BIPOC cmty has fought for years..	⑩
Reply	Abortion	(-) .. Christians uphold Biblical values and honor scripture. How can you support #abortion and #homosexuality as a democrat and also be a sincere #Christian..	⑪
		.. I support the right to an abortion, as does the Kansas constitution as interpreted by the KS Supreme Court .. I say reproductive freedoms because there is more to it than just abortion services	⑫
	Capitol	.. Today, men, women, young people, and children peacefully protested at their US Capitol. Only a small few broke the law. #StopTheSteal	⑬
		(-) Weird, youd think a Patriot and veteran would have stopped people from attacking the nations Capitol right instead of running away ..	⑭
	Incivility	(+) .. I have an issue with stupid rules that require mask wearing when no one is present .. so if they choose to have dumb rules that's their right. It's also my right to criticize them	⑮
		.. Would be threaten Dr Fauci- trumpsters, militia types - all those who misunderstand the meaning of freedom; the greater good Morons	⑯
	Anger	(+) .. by potential criminals & untrained adults are you referencing Hunter Biden? Because surprisingly you haven't appeared to tweet about his lies on a federal form to obtain a gun which was later left in a public trash can	⑰
		.. Maybe yall should consider not spreading lies and conspiracy theories then you might get to play with grownup toys #GOPSeditiousTraitors #Copolitics	⑱
	Anxiety	(+) .. Current Missouri law outlaws the mere possession of brass knuckles, which is exactly the kind of thing many of us are worried some would do with guns. I don't see how your analogy applies	⑲
		Donald Trump is a danger to our country. He's unhinged and I personally fear what harm he has planned for Inauguration Day. We need to feel safe before we can heal #TrumpIsDANGEROUS	⑳
Mention	Abortion	(-) Even if the Texas abortion law is bad politics, it has inarguably already saved lives .. I'd rather good policies that save lives than good politics that save Republicans..	㉑
		.. People should make their own decisions about abortion without involving the government	㉒
	Rights	.. Blaming a president or vice president? .. Demodupes keep getting angry that they're getting their rights trampled on by Democrat politicians, but refuse to vote out the culprits.	㉓
		(+) .. ending taxpayer subsidies for a foreign government that commits human rights abuses against millions of people is just so far beyond..	㉔

Table 5: Examples of content characteristics related to the visibility of cross-cutting posts between Republicans and Democrats (R→D) and Democrats and Republicans (D→R) are provided. Features associated with higher visibility are marked with (+), while those linked to lower visibility are marked with (-). *Note: This table is illustrative rather than exhaustive and highlights representative content characteristics that are most strongly associated with visibility differences (RQ1b).*

Interaction	Feature	Examples	No.
Mention	Gun	■ Don't let progressive liberal @xx take away your constitutional right to protect your loved ones..Gun sales in major swing states up nearly 80% this year: Will it have any bearing on election outcome..	(25)
		■ (+) .. I am not embarrassed and will not apologize for receiving support from families and survivors of gun violence. This happened to my community and we are still dealing with the consequences..	(26)
	Incivility	■ (-) .. Leftists live in an alternate reality which I refer to as leftist lunacy land where the Dems won the civil war. Revisionist history as they try to rewrite all history in their favor..	(27)
		■ (-) .. Talk about Trump; my God the world is endingTrumps a demagogue, ignorant, idiotic; egotistical. GOP cares only about power, not the truth	(28)
	Generalize	■ (-) .. #Patriots - stand tall and keep praying, tell all your friends. It will be fair, legal and legitimate or it will be nothing at all..	(29)
		■ (+) .. the last thing Donald Trump needs in the world is this job President. Lets help him out and #vote #BidenHarris #November3rd	(30)
	Causality	■ (-) .. Nearly three dozen people died in Ontario because coronavirus policies delayed their heart surgeries..	(31)
		■ .. They helped plan and bring terror and war to us and many other nations. Just sayin. Not like Taliban Afghanistan was just minding its own business and we went there..	(32)
	Subjectivity	■ (+) .. is the most liberal Republican representative Wyoming has EVER sent to D.C. Embarrassing!!! XX needs to resign.	(33)
		■ (+) .. This could have something to do with our Top Ten Ranking in New Cases of COVID.	(34)
	Hedges	■ (-) .. Pressure mounts on XX to identify funders, organizers of violent riots — Just The News..	(35)
		■ .. I hope XX gets a season in this year, but if even Jose Canseco says you're a 'fool', you gotta wonder if your plan makes sense..	(36)

Table 6: **Table 5 (continued).** Examples of content characteristics related to the visibility of cross-cutting posts.

Posting style	%	Republican example	Democrat example
Subjectivity	52.5	*I'm game if we can also require ID and confirm the actual holder of said ballot is actually the person casting the vote and that they are eligible to vote! I think we could all meet in the middle on this! Great opinion piece below..*	*..California can now sanction MediCal health care plans that don't comply with screening and testing the children most at risk of lead poisoning..*
Causality	35.1	*..Why don't you look up the context of Ms. King's quote? It was made in reference to McCarthy's remarks about CRT; institutional racism. When I quoted Dr. King it was SOLELY in reference to the assertion that one is inherently racist based on their skin color. A big difference*	*PFAS disposal is really just another step in the contamination cycle. Feeding the Waste Cycle: How PFAS Disposal Perpetuates Contamination..*
Hedges	37.1	*..'I feel very nervous' about Biden's chances against Trump..*	*..I wish we all realized that such training is not about blame or finger point though perhaps uncomfortable at times, same as students often are as they learn, but we ask them to keep learning. If they can, we can. Hope we can inspire one another.*
Generalization	17.3	*We must focus on STEM and rid ourselves of CRT..China Rises as Worlds STEM Leader While American Schools Place Diversity First..*	*Once again, XX reports the truth about NY's inability to execute. This time it's administering COVID vaccinations..*
Positive emotion	57.6	*Poll: Voters Favor Trump over Biden in Hypothetical 2024 Matchup..*	*While XX is focused on bringing students safely back into the classroom, there are families who feel safer with at-home learning options. We will urge XX to make sure schools have the resources to expand online learning options for students.*
Anger	25.1	*VICIOUS: Grown Woman Assaults 12-Year-Old Boy in Denver Over Pro-Trump Yard Sign..*	*Ignorance now rules America. Not the simple, if somewhat innocent ignorance that comes from an absence of knowledge, but a malicious ignorance forged in the arrogance of refusing to think hard about an issue, to engage language in the pursuit of justice..*
Anxiety	20.3	*While XX frantically repeats lies about the admins handling of the border, federal authorities are releasing thousands of Covid positive illegal aliens into the U.S..*	*..And it was really critical here. Things will not be the same for large and small districts. Local as possible would have offered the least disruption.*

Notes. Posting styles are binary indicators (present/absent) as defined in Section 4.1. Examples are shortened and de-identified; they are intended to illustrate what each construct captures.

Table 7: Prevalence of posting styles and representative examples by party. Percentages are computed over all posts containing at least one interaction (reply, mention, or retweet) in the estimable sample. Examples are illustrative and anonymized.

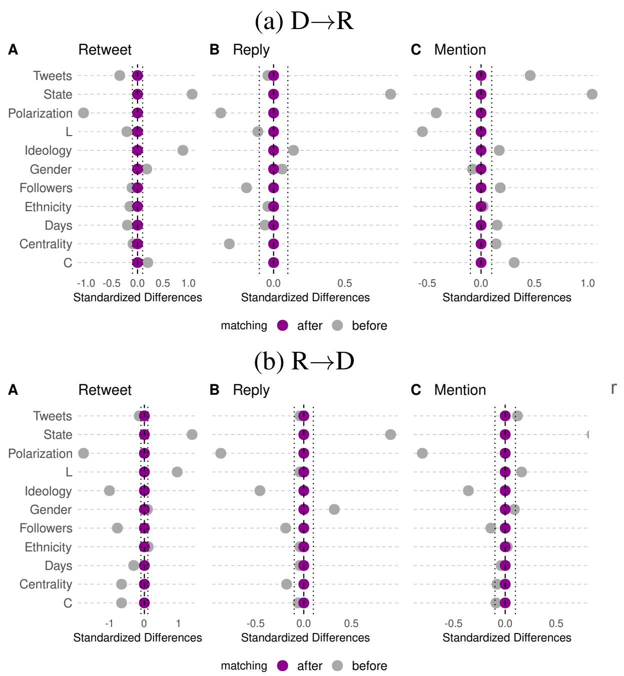

Figure 11: **Covariate balance for matching.** The standardized differences for each of the covariates before and after matching for cross-cutting posts by (a) Democrats and (b) Republicans. All the covariates are balanced (i.e., score between -0.1 to 0.1) after matching in each case.

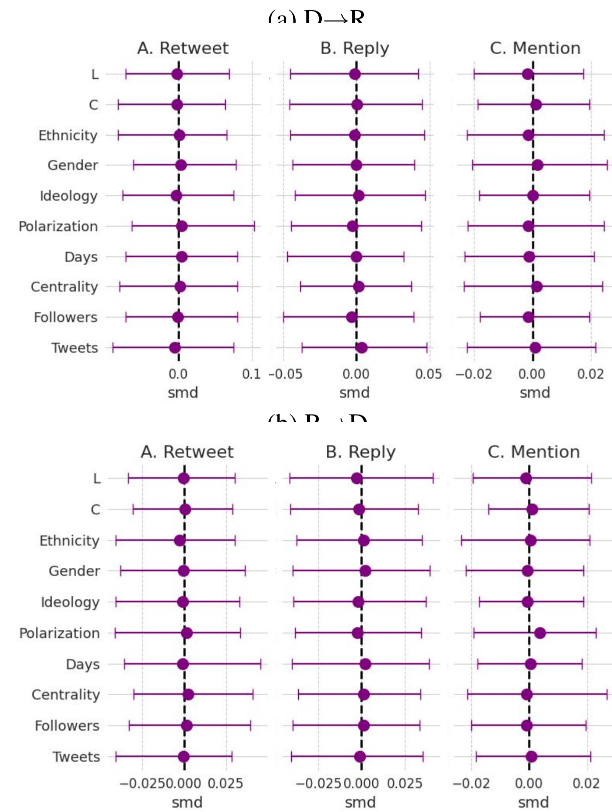

Figure 12: **Distribution of Standardized Differences for Covariates.** The mean and 95% confidence intervals of standardized mean differences (smd) for covariates after matching for cross-cutting posts by (a) Democrats and (b) Republicans. The feature 'State' is not shown in the analysis due to its sparsity, which leads to multicollinearity issues in the calculation of standardized differences.

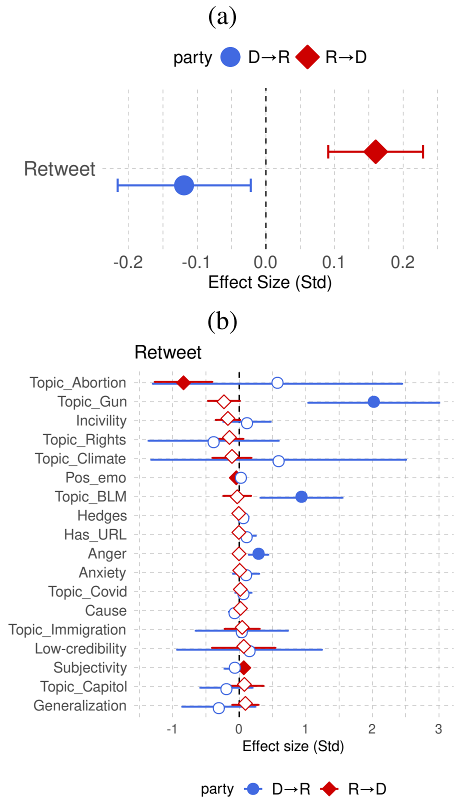

Figure 13: Effect of (a) cross-talk (RQ1a) and (b) corresponding posting styles (RQ1b) on engagement for Retweet.

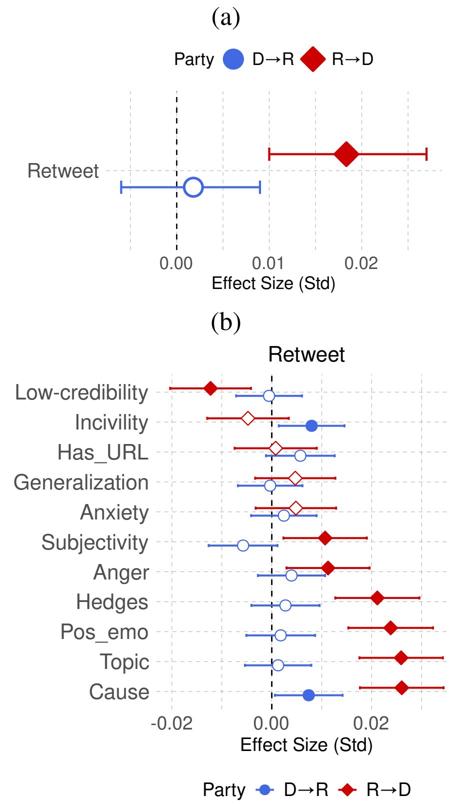

Figure 14: Effect of engagement on future (a) cross-talk rate (RQ2) and (b) stylistic choices (RQ3) for Retweet.

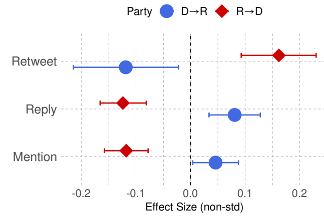

Figure 15: Non-standardized effect sizes of cross-cutting interactions on engagement (RQ1a).

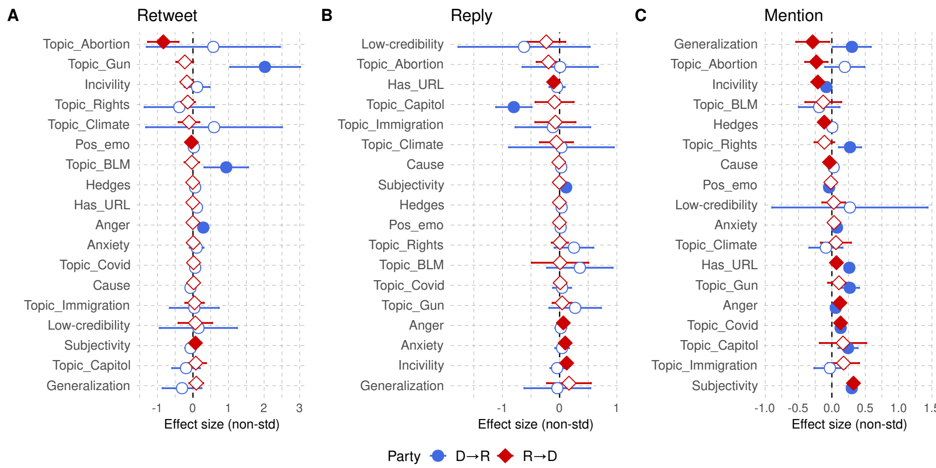

Figure 16: Non-standardized effect sizes of posting styles on engagement of cross-cutting interactions (RQ1b).

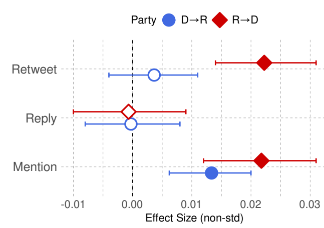

Figure 17: Non-standardized effect sizes of engagement on future rate of cross-cutting interactions (RQ2).

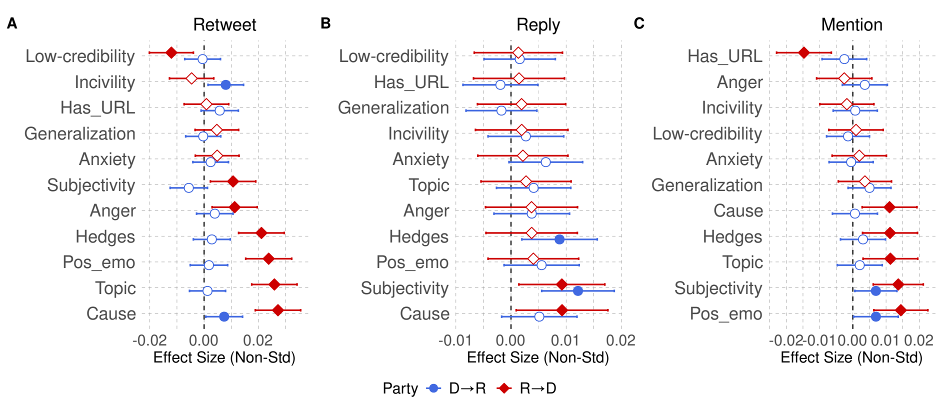

Figure 18: Non-standardized effect sizes of engagement on rhetoric and style of future cross-cutting interactions (RQ3).