Blameocracy: Causal Rhetoric in Politics*

Francesco Bilotta Alberto Binetti Giacomo Manferdini
(JMP)

March 25, 2026

Abstract

This paper studies the supply and effects of causal rhetoric in U.S. politics. We define causal rhetoric as assigning responsibility for political outcomes, via claims of blame and merit. Training a supervised classifier, we detect causal rhetoric in over a decade of congressional tweets, finding that its supply has risen rapidly and pervasively, displacing affective messaging. We show that the production of causal rhetoric involves a trade-off between revenues and costs. First, quasi-random variation in Twitter adoption shows that blame increases small-donor revenues by expanding donor count, while merit raises average donation size. Second, fine-grained legislative data suggest that policy ownership determines relative costs: blame is cheaper for opponents, merit for proposers. Finally, causal rhetoric has downstream effects on societal outcomes, fostering protest activity and relating to polarization and institutional trust.

Keywords: Social Media, Narratives, Text-as-Data, Campaign Finance, Elite Polarization, Protests

---

*Francesco Bilotta: Bocconi University, francesco.bilotta2@phd.unibocconi.it; Alberto Binetti: Princeton University, abinetti@princeton.edu; Giacomo Manferdini: Bocconi University; giacomo.manferdini@phd.unibocconi.it. We are grateful to Enrico D. Turri for his input in earlier stages of this project. We thank Luca Braghieri, Francesco Capozza, Sarah Eichmeyer, Nicola Gennaioli, Gloria Gennaro, Matteo Grigoletto, Dirk Hovy, Rafael Jiménez-Durán, Tania Lombrozo, Massimo Morelli, Jaime Marques Pereira, Egon Tripodi, Carlo Schwarz, and Ekaterina Zhuravskaya for their helpful comments and suggestions, as well as participants to Munich Workshop on Beliefs Narratives and Memory, 2nd Bocconi-CCA-Cornell Political Economy Workshop, Princeton Junior PE Seminar, Workshop on Text-as-Data in Economics at Lancaster, 2nd Verona Early Career Workshop in Economics, Potsdam Text-as-Data in Behavioral Economics Workshop, XIV Alghero IBEO Workshop, 10th Monash-Paris-Warwick-Zurich-CEPR Text-As-Data Workshop, 1st CEPR Future of Democracy Conference at WZB, Ludwig Erhard ifo Research Workshop, and 18th UniTo-CCA PhD Workshop in Economics.

"To err is human. To blame someone else is politics."

Hubert H. Humphrey, Vice President of the United States (1965-1969)

Political actors invest heavily in communication with voters. In the 2024 U.S. electoral campaign, total spending reached a record $15 billion, with nearly half allocated to media operations (including consultancy) and an additional 5% to quantitative research specifically (OpenSecrets, 2025). This funding allocation reveals a firm belief that the content and structure of political communication can shape voter behavior. The content dimension is well understood: rhetoric aims to improve a candidate's relative standing, typically through attacks on opponents and, conversely, self-promotion (e.g., Lau and Rovner, 2009). What about structure?

A growing body of research in behavioral and political economics suggests that causal framing may be a particularly effective tool of persuasion. When information is embedded in explanations, it is more likely to shape beliefs (Alesina et al., 2023) and decisions (Hüning et al., 2022), to be sought out (Bursztyn et al., 2023), remembered (Graeber et al., 2022), and transmitted (Graeber et al., 2024).[1] While this literature highlights the persuasive potential of causal language, we know very little about how politicians use causal rhetoric in practice; whether it delivers the political returns they seek; whether its use comes with strategic costs; and how it shapes voters' offline behavior and attitudes. Our paper aims to answer these questions.

We define causal rhetoric as assigning responsibility for political outcomes, via claims of blame and merit. Through supervised classification, we detect it in a large corpus of tweets (4.2 M) posted by U.S. Members of Congress between 2012 and 2023.[2] To start, we document that the supply of causal rhetoric rises rapidly and pervasively over our sample period, increasingly displacing purely affective messaging. We interpret this shift through the lens of production theory, analyzing both revenues and costs of causal rhetoric. Leveraging quasi-random variation in early Twitter adoption, we show that blame increases small-donor revenues by expanding the number of donors, while merit raises the average donation size. Using fine-grained data on legislative activity, we provide evidence that rhetorical choices are constrained by policy ownership, making opponents of a bill more likely to shift blame, while proposers are more likely to claim merit. Finally, we show that causal rhetoric impacts societal outcomes: blame increases incidence of protests, while merit increases their the number; blame is associated with lower trust in government and greater affective polarization – with the opposite pattern for merit.

The primary challenge we face is measurement. Standard tools in computational linguistics struggle to detect causality because causal cues are often implicit rather than

[1] More broadly, cognitive psychology shows that human reasoning centers on causal inference (see Chater and Loewenstein, 2016; Lombrozo and Vasilyeva, 2017; Sloman and Lagnado, 2015, for reviews).

[2] Our dataset approximates the universe of tweets from House Representatives over 2013-2023, and that of tweets from Senators, restricting to 2017-2023.

signaled by fixed syntactic or semantic markers.[3] We address this limitation using a supervised learning approach based on bidimensional classification. A tweet is coded as causal if it attributes a potential outcome to the (hypothetical) intervention of a political agent. Separately, we assign each tweet a tone – positive, negative, or neutral – based on the attitude expressed toward its subject.[4] Within causal tweets, those with positive tone are labeled as merit, and those with negative tone as blame. We refer to this classification as the tweet's rhetorical style.

Based on this definition, we hand-label a training set of approximately 4,000 tweets, obtaining high inter-annotator agreement (Fleiss' Kappa = 0.64). We then fine-tune a RoBERTa-large model (Loureiro et al., 2022) – pre-trained on 154 million tweets – to classify tweets as expressing merit, blame, or none. We define a tweet as causal, ex-post, if its predicted style is merit *or* blame.[5] The model achieves strong performance: 0.83 accuracy, 0.84 F1-score, and 0.73 Matthews Correlation Coefficient – comparable to or exceeding standard benchmarks in the literature.

We validate the resulting measure through a series of internal and external checks. Intuitively, blame texts are predominantly directed at others or out-groups, while merit texts refer to the self or in-group – both syntactically, via pronoun use, and semantically, using targets identified by the Political DEBATE language model (Burnham et al., 2024). Consistent with psychological theories of responsibility (Malle et al., 2014), blame tweets are more retrospective, while merit tweets are more prospective. These patterns are reinforced by diagnostic bigram analysis. For external validation, we correlate our labels with independent annotations of credit-claiming and policy-attack statements from America's Political Pulse (Westwood et al., 2024) and find strong correlations at the politician level. Finally, we show that our blame-merit measure is largely orthogonal to sentiment, emotionality, and moral rhetoric (Enke, 2020; Gennaro and Ash, 2022; Hutto and Gilbert, 2014), indicating that it captures a novel rhetorical dimension.

Our conceptual innovation enables us to document three key facts about the supply of causal rhetoric in congressional communication. First, causal rhetoric has become widespread over our sample period. Blame and merit tweets account for 19 percent of congressional tweets in 2012 (with both dimensions starting at around 10 percent) but rise to 43 percent in 2023 (both converging at around 20 percent). The increase is steepest between 2017 and 2019, with stable levels before and after, suggesting a structural shift rather than a temporary shock.

Second, the rise is pervasive. An event-study design shows that the increase persists

[3]For instance, "causality" is one of the worst-performing labels in LIWC-22 (Pennebaker et al., 2022), a gold standard for dictionary-based methods.

[4]We refer to this as tone rather than sentiment, as our annotation captures evaluative nuance that standard sentiment dictionaries like VADER (Hutto and Gilbert, 2014) often miss.

[5]This approach is supported by the empirical rarity of neutral-toned causal tweets in our training data. We also validate this decision by training an independent classifier for causal vs. non-causal language and find a high correlation with the synthetic label.

after controlling for both politician and topic fixed effects: accounting for composition absorbs less than one fourth of the rise in blame and inflates the rise in merit by about one fifth. The increase is broadly distributed across politicians, but disproportionately concentrated in policy domains, amplifying pre-existing topic-level differences.

Third, as causal rhetoric rises, it crowds out purely affective messaging – defined as attacks or self-promotion based only on tone. The share of blame among negative tweets rises from 23 to 42 percent; the share of merit among positive tweets from 13 to 31 percent. As a result, elite-level polarization, which was initially equally conveyed across causal and non-causal tweets, becomes almost exclusively concentrated in blame and merit by the end of the period.[6]

To understand the economic forces driving the supply of causal rhetoric, we draw an analogy between tweet posting and a production problem, in the spirit of "price theory", following Aridor et al. (2024).

Our first step is to quantify the returns to producing blame or merit tweets. We focus on campaign contributions from small donors, defined as donations below $1,000. This choice – standard in the literature (e.g., Petrova et al., 2021) – offers several advantages, including that small donations serve as a proxy for broader political support, beyond their monetary value.[7] To address identification, we leverage quasi-exogenous variation in Twitter penetration induced by the platform's early diffusion following the South-by-Southwest (SXSW) festival in 2007. Specifically, we instrument the county-level Twitter users with the number of followers of the official SXSW Twitter account in 2007 (Fujiwara et al., 2024; Müller and Schwarz, 2023).

We find that blame increases aggregate revenue from small donations. In particular, a one standard deviation increase in the share of blame tweets raises revenue from donations by 3.4 percent in the average county. Decomposing the effect reveals that blame operates at the extensive margin, through a mobilization channel. It increases the number of donors by 2.2 percent in the average county, while it has no significant effect on average donation size. Consistent with this interpretation, we show that blame spreads virally: whereas blame accounts for only 15 percent of tweets, it generates nearly 40 percent of all retweets. Finally, donor-level ideological heterogeneity reveals that extreme donors respond strongly to blame, while moderates do not.

Merit, by contrast, has no significant effect on aggregate revenues and shows mirror-like regularities compared to blame. First, merit impacts the intensive margin, through a fidelization channel: a one standard deviation increase raises the average donation by about 0.8 percent in the average county. Second, moderates respond positively to merit,

[6]Our measure of polarization is the difference in sentiment between the tweets posted by the ruling and the opposing party, aggregating at the presidency level.

[7]Compared to large or PAC-style contributions, small donations are less likely to reflect lobbying or access-seeking motives, are more likely to capture expressive intent, and come from individuals who are more representative of the general U.S. population (Bouton et al., 2022).

whereas extreme donors respond negatively.

Taken together, these findings suggest that while blame delivers higher average returns, blame and merit function as complements – each serving distinct strategic purposes and appealing to different constituencies. Then, just as the usage shares of complementary inputs reveal their relative prices, the allocation of a politician's tweets between blame and merit offers insight into the underlying costs of each strategy. We next leverage this analogy to identify the cost structure associated with blame and merit.

Comparing the share of merit and blame tweets for each member of Congress, we find that a clear trade-off emerges once causal rhetoric becomes widespread. Politicians in opposition tend to resolve this trade-off in favor of blame, while those in government favor merit; no fixed trait – such as demographics or ideology – predicts this pattern. This suggests an underlying reputational cost: blaming others is intrinsically less credible when in power, while claiming credit is less credible when out of office.[8] To support this mechanism more directly, we use bill introductions as a proxy for observable political action – providing a more fine-grained measure than power status alone. We find that when the opposing party introduces a bill, blame increases and merit decreases; the pattern reverses when the bill comes from one's own party, confirming our mechanism. Notably, similar to what observed for the tradeoff, these effects become pronounced only after the widespread diffusion of causal rhetoric – especially for blame.

Whereas our analysis primarily focuses on the production of causal rhetoric, a large literature on the societal effects of social media (Campante et al., 2022; Zhuravskaya et al., 2020, for reviews) suggests that exposure to persuasive content can have unintended consequences for offline political behavior. We conclude by exploring two leading outcomes: protest activity and voters' attitudes toward peers and government.

Adapting our geography-based design to protest outcomes, we find patterns consistent with the mobilization and fidelization channels. A one standard deviation increase in the share of blame tweets posted by all politicians raises the likelihood of a protest by nearly 10.1 percentage points in the average county. In contrast, a one standard deviation increase in the share of merit tweets posted by all politicians is associated with a 22.6 percent rise in the number of protests in the average county.

Turning to political attitudes, we exploit survey data from Westwood et al. (2024) to correlate state-level exposure to rhetorical styles with voter beliefs. We find that blame is positively associated with affective polarization and linked to lower trust in government and reduced perceptions of government responsiveness. Merit, by contrast, is associated with more positive attitudes across all dimensions.

[8] Bilotta and Manferdini (2024) formalize a similar constraint in a model of narratives based on voters' partial identification of policy effectiveness (cf. Manski, 1995).

Related Literature

We contribute to various strands of the literature, discussed thematically below.

Social Media. We contribute to the growing literature on the economics of social media (for a recent review, see Aridor et al., 2024). Closest to our work is a recent strand that studies the effects of Twitter on campaign contributions. Petrova et al. (2021) show that politicians benefit from adopting Twitter, especially entrants and in high-penetration states; Boken et al. (2023) identify a fundraising premium associated with tweets "going viral"; and Rotesi (2019) show that Twitter diffusion increases donations from Republican voters.[9] Relative to this work, we shift the focus from platform adoption and diffusion to the persuasion strategies employed by politicians. We quantify both the benefits and costs of a specific rhetorical form – causal attribution – by analyzing the content and structure of political messages, a dimension largely neglected in the existing literature. In doing so, our work bridges the gap with the empirical literature on persuasion, which has extensively studied how traditional media influence voters (e.g., DellaVigna and Gentzkow, 2010; DellaVigna and Kaplan, 2007; Enikolopov et al., 2011).

Similarly, we contribute to the strand of the literature on the societal effects of social media (for reviews, see Campante et al., 2022; Zhuravskaya et al., 2020), particularly to work on protests (e.g., Boyer et al., 2024; Enikolopov et al., 2020; Gylfason, 2023) and polarization (e.g., Allcott et al., 2020). We add to this line of research by showing that beyond content and sentiment slant, the causal framing of tweets plays a role in shaping political behavior and attitudes.

Text-as-Data Our paper also contributes to the literature on text-as-data in political economy (see Ash and Hansen, 2023; Gentzkow et al., 2019, for reviews). Conceptually, we introduce a novel definition of causal attribution in politics – specifically, merit-taking and blame-shifting – and measure its prevalence in elite political communication. In this sense, we add to prior work that quantifies other linguistic dimensions of political rhetoric, such as emotionality (Gennaro and Ash, 2022), moral terminology (Enke, 2020), and linguistic complexity (Di Tella et al., 2023).

A closely related contemporary contribution is Algan et al. (2025), which documents a sharp rise in negative emotions – particularly anger – in political tweets after 2016. Their findings complement ours in both timing and substance: the emotional shift they document aligns with the post-2016 surge in causal rhetoric, and the prominence of anger – strongly associated with blame, as we show below – offers a psychological backing for

[9]These papers differ in sample periods and strategies: Petrova et al. (2021) restrict to 2009-2014 and analyze heterogeneity by candidate status; Boken et al. (2023) focus on 2019-2020 and exploit SXSW as an instrument; Rotesi (2019) focus on election years and instrument Twitter exposure with the relocation of NBA players active on the platform.

the patterns we observe.

Narratives. Finally, we contribute to a literature studying persuasion through narratives (see Barron and Fries, 2024b, for a review), providing a field-level measurement of merit and blame attribution (see Bilotta and Manferdini, 2024; Eliaz and Rubinstein, 2025; Eliaz et al., 2023, for models formalizing this idea). In this sense, we add to a small set of papers using observational data to measure narratives in natural settings (e.g., Gehring and Grigoletto, 2023; Goetzmann et al., 2022; Macaulay and Song, 2023).[10] Relative to this strand of papers, we differ in several respects. First, our approach is domain-agnostic, whereas most existing studies focus on specific topics or events. Second, we capture both the extent and direction of causal attribution without imposing structure on how causality is expressed or which agents are involved – relying instead on the semantic content of the text itself.

1 Data

Our analysis combines data from multiple sources covering political communication on social media, campaign donations, legislative activity, and protest activity, enriched with demographic and ideological information on elected officials.

Twitter Data Our primary dataset consists of approximately 4.2 million tweets posted by Democratic and Republican members of the U.S. Congress between 2012 and 2023. We integrate data from the CongressTweets project (CongressTweets, 2023) and Bellodi et al. (2023). To the best of our knowledge, this dataset includes all tweets posted by House members from January 3, 2013, to July 11, 2023; all tweets by House incumbents running for re-election in 2012 (covering the full calendar year); and all tweets by Senators from June 21, 2017, to July 11, 2023. In total, the sample comprises 4,198,455 tweets from 900 unique politicians across 1,789 unique Twitter accounts.[11] We enrich these data with demographic and ideological information from ProPublica, VoteView, and Wikipedia. Descriptive statistics at the tweet and politician levels are provided in Table A1. While Republicans represent a slight majority of politicians in our sample, the majority of tweets are authored by Democrats, who – as also shown by Fujiwara et al. (2024) – tend to be more active on Twitter. Most politicians in the dataset hold at least a bachelors degree, the average age is slightly under 60, and approximately one quarter are female.

[10] Most of the existing empirical work on narratives relies on surveys (e.g., Andre et al., 2021) or experiments (e.g., Barron and Fries, 2024a; Kendall and Charles, 2023) instead.

[11] We exclude retweets, as they do not represent original content. Quoted tweets are retained because they also contain user-generated text.

Donations We collect information about campaign contributions from the Database on Ideology, Money in Politics, and Elections compiled by Bonica (2024). This dataset includes over 850 million donations made by individuals and organizations to candidates in local, state, and federal elections between 1979 and 2024. We restrict our analysis to small donations by individuals directed to candidates for the U.S. House of Representatives and U.S. Senate from the 2012 through the 2024 election cycles. In particular, our definition of small donations includes those under $1,000. This results in a final sample of approximately 211 million unique donations.

Bills We also construct a dataset of congressional bills using official records from the U.S. Government Publishing Office, accessed via the GovInfo bulk data repository. We scrape structured metadata for all bills[12] introduced in the U.S. House and Senate from the 112th Congress (2011-2013) through the 118th Congress (2023-2025). This yields approximately 100,000 distinct legislative items. In an attempt to focus on politically relevant legislation, we restrict our attention to bills sponsored by Democratic or Republican legislators that received at least one roll call vote. The final sample consists of 3,182 bills sponsored by 754 unique legislators between 2012 and 2024.

Protests We incorporate data on protest activity from the Crowd Counting Consortium, hosted by the Ash Center for Democratic Governance and Innovation at Harvard University (Crowd Counting Consortium, 2025). The dataset documents protest events across the United States, spanning a wide range of political and social causes. We use data from the first two phases of the project, covering the periods 2017-2020 and 2021-2024, respectively. In total, the dataset includes 212,004 recorded protest events.

America's Political Pulse We use data from America's Political Pulse, a project led by the Polarization Research Lab that tracks both elite political rhetoric and public attitudes in the United States (Westwood et al., 2024). On the elite side, the project classifies congressional communications – across speeches, newsletters, tweets, press releases, and public statements – into rhetorical categories. These measures are available at the individual politician level starting in 2023. We use all observations available through the end of our sample period, covering 359,093 communications from 456 legislators. On the voter side, the project fielded a weekly nationally representative survey since September 2022, measuring attitudes such as affective polarization, trust in politicians, and perceptions of government responsiveness. Focusing on answers until September 2023 leaves us with 59,228 observations.

[12]Specifically, we include House bills, House joint resolutions, House concurrent resolutions, and House simple resolutions, as well as their Senate counterparts.

2 Measuring Blame and Merit in Text

The cornerstone of our analysis is a bidimensional classification that establishes whether a tweet attributes outcomes to political agents, and, if so, whether the attribution is negative (blame), positive (merit), or neither. To construct this measure at scale, we train a supervised classifier based on a RoBERTa-large model (Loureiro et al., 2022). An array of validation exercises shows that the classifier is accurate and that the resulting measure captures a rhetorical dimension distinct from those already studied in the literature.

2.1 Definition of Blame and Merit

Our classification rests on two dimensions: *causality* and *tone*.

First, causality is binary. A tweet is causal if it attributes a potential outcome to the power status of a political agent. Political agents include politicians, institutions, and politically aligned organizations, but exclude natural events (e.g., pandemics, disasters) and neutral actors (e.g., scientific teams). Power status refers to an agent's capacity for meaningful political action, either factual (currently or previously taken) or hypothetical (could be taken if in power). Potential outcomes may concern any social (e.g., civil liberties) or economic consequence (e.g., GDP growth). Importantly, causal attribution does not require explicit connectors such as "because" or "since."[13]

Second, tone captures the stance of a tweet toward its subject. We classify tone as negative (-1), neutral (0), or positive (1). Negative (positive) tone reflects unfavorable (favorable) language; neutral tone reflects descriptive statements without evaluative language. Our manual annotation captures nuances that dictionary-based methods may instead miss.[14]

Finally, combining causality and tone yields a synthetic measure of *rhetorical style*. Tweets that are causal and negative are coded as blame (-1); tweets that are causal and positive are coded as merit (+1); all others are coded as none (0). Table A2 provides labeled examples.

2.2 Classification Pipeline

We next describe the pipeline that allows us to implement our measure of rhetorical style at scale.

First, we construct a labeled dataset of 3,958 tweets. A reasonable prior is to expect that blame and merit tweets represent a minority of the overall corpus. At the same time, balanced representation across classes is a crucial element to ensure proper training of

[13]For example, "Politician X is corrupt" is evaluative but not causal, while "Politician X's corruption undermines democracy" is causal under our definition.

[14]For instance, "Our policies avoided a tragedy" would register as negative sentiment in a dictionary method, but has a positive tone under our coding.

the classifier. To address this, we aim to increase their representation through targeted oversampling, which is a common approach in the literature (e.g., Davidson et al., 2017; He and Garcia, 2009; Talat and Hovy, 2016). Hence, we adopt a two-step sampling strategy. We begin by generating benchmark examples of merit and blame tweets by Democrats and Republicans using ChatGPT 4o. We then compute cosine similarities between these examples and the main corpus with SBERT-mini embeddings (Reimers and Gurevych, 2019). Half of the tweets selected for annotation are those most similar to the benchmarks (balanced across party and style), while the other half are drawn at random. In addition, 530 tweets are jointly annotated by three coders to assess inter-annotator reliability. For this subsample of tweets jointly annotated, labels for causality and tone are assigned by majority vote, with ties broken randomly. Our protocol yields an average pairwise correlation of 0.73 and a Fleiss Kappa of 0.64 – generally interpreted as substantial agreement (the second best category out of five). Coders agree on rhetorical style in 67 percent of tweets, roughly six times the rate expected by chance.

Second, we fine-tune a RoBERTa-large model pre-trained on 154 million tweets (Loureiro et al., 2022).[15] We split the labeled corpus into 80 percent training data (3,166 tweets) and 20 percent validation data (792 tweets).[16]

Finally, the classifier assigns each tweet a probability distribution over the three classes – merit, blame, and none – and classifies the tweet to the highest-probability class. Ambiguous cases, where the model assigns similar probabilities to multiple categories, are extremely rare.[17] In our corpus, 20 percent of tweets are coded as merit, 16 percent as blame, and the rest as none. On the validation set, the model achieves an accuracy of 0.83, an F1-score of 0.84, and a Matthews Correlation Coefficient of 0.73, demonstrating strong performance.

Our analysis relies on two main variables: an indicator for blame and an indicator for merit. Since neutral-toned causal tweets are exceedingly rare in our hand-labeled training data – accounting for just 0.4% of cases – we use the term causal to refer to tweets that contain either blame or merit. To validate this choice, we train a dedicated classifier using only the "causal" label from the annotated dataset and find that its output is highly correlated with our combined blame-or-merit indicator ($\rho = 0.9$).

[15]Fine-tuning adapts a pre-trained language model to a specific downstream task by adjusting its parameters on a smaller labeled dataset. In our context, this procedure improves the model's ability to capture causal attributions and tone in political discourse.

[16]We train for 10 epochs, selecting the epoch with the highest F1 score. As robustness checks, we also fine-tune (i) a version of the model with half of its layers frozen and (ii) the BERTweet model (Nguyen et al., 2020). Both yield marginally lower performance but highly correlated outputs, with Matthews Correlation Coefficients of 0.93 and 0.85, respectively.

[17]Figure A1 shows that, conditional on selecting a certain label, the probability distributions are strongly left skewed and concentrated close to 1, indicating that the classifier is rarely uncertain.

2.3 Validation

We next validate our measure, showing that blame and merit texts display linguistic and semantic patterns consistent with an intuitive understanding of these concepts.

First, we examine the target of causal attributions. Syntactically, we measure whether tweets are self- or other-referential by comparing the relative frequency of second- and third-person pronouns versus first-person pronouns. Figure A2a shows that blame tweets are directed outward (i.e., framed around others) whereas merit tweets focus more on the self. Semantically, we classify whether the causal attribution targets Democrats or Republicans using the Political DEBATE language models (Burnham et al., 2024).[18] Restricting to the 40 percent of tweets where party identity can be recognized, Figure A2b shows that blame is disproportionately aimed at the opposing party, while merit is more often directed inward.

Second, we turn to the temporal dimension. Causal arguments can be retrospective – attributing responsibility for past outcomes – or prospective – linking current actions to future consequences. We capture this using normalized counts of past versus future tense. Figure A2c shows that blame is predominantly retrospective, while merit is more forward-looking, consistent with findings from psychology (Malle et al., 2014).

Third, we examine which emotions are conveyed by each rhetorical type. Using a RoBERTa model fine-tuned for emotion detection in tweets (Camacho-Collados et al., 2022), we disaggregate the emotional content of blame and merit tweets. Figure A6 shows that blame is dominated by anger (nearly 50%), followed by disgust and fear – consistent with the idea that anger is directional and linked to causal attribution (Lazarus, 1991). Merit, by contrast, is overwhelmingly associated with optimism (over 50%), reinforcing our earlier finding that merit is forward-looking.

Fourth, we study diagnostic language. Following Gentzkow and Shapiro (2010), we extract the bigrams most distinctive of each category. As shown in Figures A3 and A4, the bigrams align with intuition and the earlier patterns.[19] We complement these diagnostics with illustrative tweets in Tables A3 and A4.

Finally, for external validation, we compare our blame and merit indicators with measures from the America's Political Pulse dataset, which tracks politicians' credit-claiming and policy-attack statements across multiple communication channels beginning in 2023 (Westwood et al., 2024).[20] For each politician, we compute the share of credit-claiming

[18]Political DEBATE is a language model specialized in zero-shot and few-shot classification of political documents, with performance on par with, or better than, state-of-the-art language models.

[19]For Democrats, "trump administration" is diagnostic of blame, while "act will" signals merit. For Republicans, "southern border" characterizes blame, while "act will" again identifies merit.

[20]In their codebook, Westwood et al. (2024) define credit-claiming as *"communications about creating or passing legislation; securing government spending, grants, or funding; or emphasizing personal or party accomplishments in office"*, and policy-attack as *"communications about objecting to or raising concerns about a specific policy, law, or court ruling; using fact-based arguments even if critical or negative; avoiding emotional appeals, inflammatory language, claims of extremism, or personal attacks*

and policy-attack statements and correlate these with the share of merit and blame tweets, respectively, over the same period. Figure A5 shows strong positive correlations – 0.74 between blame and policy-attack, and 0.68 between merit and credit-claiming – underscoring the external validity of our measure.

2.4 Distinctiveness of Blame and Merit Text

To conclude, we show that our measures do not reduce to a combination of well-established linguistic features of political text.

Specifically, we compare blame and merit to sentiment (Hutto and Gilbert, 2014), emotionality (Gennaro and Ash, 2022), and the prevalence of moral terminology (Enke, 2020).[21] We estimate tweet-level regressions of the form:

$$y_i = \alpha + \beta_1 \mathrm{S}_i + \beta_2 \mathrm{E}_i + \beta_3 \mathrm{M}_i + \varepsilon, \tag{1}$$

where $y_i \in \mathrm{Blame}_i, \mathrm{Merit}_i$ are binary indicators denoting whether the tweet is blame or merit, respectively, S_i is sentiment, E_i is emotionality, and M_i is moral terminology. All regressors are standardized, and errors are clustered at the politician level.

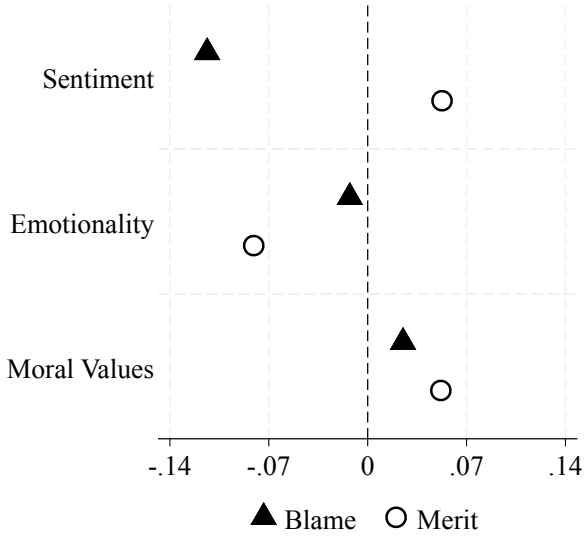
Figure 1 shows that both blame and merit are only modestly correlated with existing linguistic features such as sentiment, emotion, and moral rhetoric. The estimated coefficients are generally small – below 0.1 in absolute value – and the explanatory power of these features is limited, with R^2 values remaining under 0.1 for both dimensions. This suggests that our measures are not simply relabeling known linguistic constructs, but instead capture a distinct and previously unexplored dimension of political discourse.

At the same time, the modest correlations that do emerge are intuitive, acting as further validation. Blame is negatively associated with sentiment, consistent with its tendency to employ negative evaluative language, while, on the contrary, merit correlates positively. Both blame and merit are weakly negatively associated with emotionality, supporting the view that causal rhetoric is framed in more reasoning-oriented terms. Interestingly, among the two, merit is more negatively associated with emotionality, possibly reflecting a higher evidentiary burden when claiming credit than when assigning responsibility.

on individuals involved with the policy, including accusations of lying or withholding information".

[21] To measure emotionality and moral terminology, we follow the procedures in the respective papers. For emotionality, we embed emotional and reasoning words and compute, for each tweet, the ratio of cosine similarities with the emotional versus reasoning embedding. For moral terminology, we compute, for each moral value, the average of vice and virtue frequencies, sum across all moral values, and normalize by the number of non-stop words.

Figure 1: Correlation with Existing Text Measures



Notes: The figure presents the estimates of Equation 1. Bars represent 95 percent confidence intervals computed with standard errors clustered at the politician level.

3 The Supply of Causal Rhetoric

Using our methodology, we show that the supply of causal rhetoric has risen sharply in congressional communication on Twitter. We document three key facts. First, the share of both blame and merit tweets roughly doubles over the sample period, with most of the increase occurring between 2017 and 2020 – indicating rapid and sizable growth. Second, the rise occurs within politicians and within topics, pointing to a pervasive shift in rhetorical strategy rather than variation in the composition of speakers or salient issues of debate. Third, causal rhetoric increasingly substitutes for purely affective messaging and emerges as a primary channel for elite polarization.

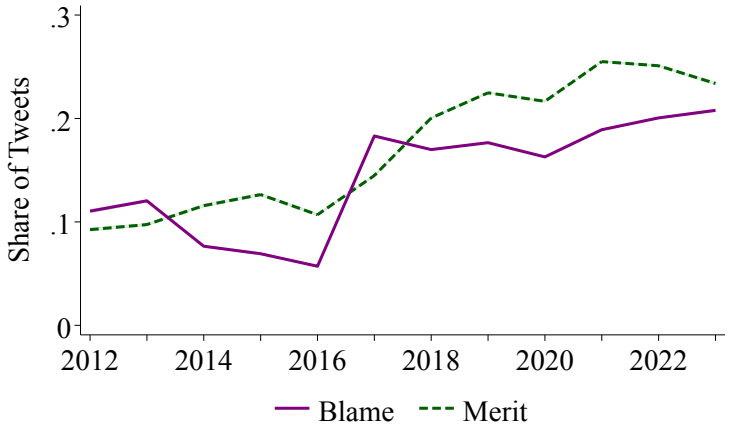
3.1 Blame and Merit over Time

Figure 2 plots the yearly share of blame and merit tweets from 2012 to 2023. Both dimensions expand markedly. Blame rises from about 10 percent to over 20 percent, while merit climbs from around 9 percent to more than 23 percent. Together, their share grows from roughly one fifth to nearly one half of all congressional tweets, underscoring that causal rhetoric has become a prevalent feature of political communication on Twitter.

The increase is also steep. The share of blame tweets increases by more than 10 percentage points in a single year, between 2016 and 2017. Merit follows a similar trajectory, but its growth is spread between 2017 and 2020. By 2020, both dimensions appear to plateau, suggesting that blame and merit became entrenched elements of congressional discourse.

Overall, these dynamics point to a structural shift in congressional communication,

Figure 2: Supply of Blame and Merit Tweets over Time



Notes: The figure presents the yearly share of tweets classified as blame and merit. Shaded areas represent 95 percent confidence intervals.

with the 2017-2020 term serving as the catalyst for the increased supply of causal rhetoric.

We expand these findings and show their robustness in different ways. First, Figure A7 replicates the analysis excluding Senators' tweets, which are missing prior to 2017, and shows very similar patterns. Second, Figure A8 shows that no other textual feature displays a comparable trend. Finally, we take some steps towards establishing external validity, showing that a correlated pattern holds for newsletters (Appendix B.1), and that the rise of causal communication cannot be explained by changes in Twitter's policies (Appendix B.5).

3.2 Politician × Topic Decomposition

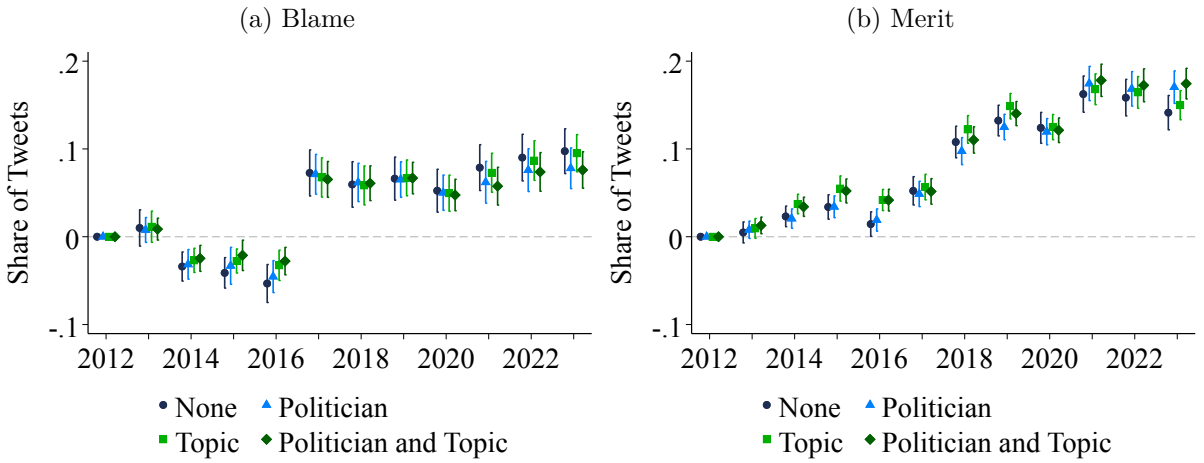
Whereas our qualitative time trends show a clear increase in the use of causal rhetoric over the past decade, this pattern alone does not speak to its pervasiveness. In principle, the rise in blame and merit could reflect a compositional shift – driven either by turnover in congressional membership or by a shift toward topics naturally prone to blame and merit. To address this possibility, we estimate an event-study specification centered on 2012, progressively controlling for politician and topic fixed effects.

We classify each tweet by topic using again the Political DEBATE language model developed by Burnham et al. (2024), assigning tweets to one of eight broad categories – economy, environment, healthcare, immigration (policy issues) and gender, gun control, policing, racial relations (sociocultural issues). We then estimate the following specification:

$$y_{iptj} = \sum_{k=2013}^{2023} \beta^k \mathbb{1}[j = k] + \lambda_p + \mu_t + \varepsilon_{iptj} \tag{2}$$

Here y_{iptj} is an indicator for blame (merit) taking value 1 if tweet i posted by politician

Figure 3: Decomposition



Notes: The figures present the estimated β^k from Equation 2. Each color corresponds to a specification including the fixed effects indicated in the legend. Bars represent 95 percent confidence intervals computed with standard errors clustered at the politician level.

p about topic t in year j is blame (merit) and 0 otherwise, while λ_p and μ_t are politician and topic fixed effects, respectively.

Figure 3 plots the estimated coefficients when gradually adding fixed effects. Consistent with earlier results, the use of blame rhetoric rises by approximately 10 percentage points between 2012 and 2023. After controlling for both politician and topic fixed effects, this increase remains substantial at 7.6 percentage points. For merit, the increase is even more pronounced: the unconditional rise of 14 percentage points expands to 17 percentage points after accounting for composition.

To complete the picture, we show that the supply of causal rhetoric is roughly homogeneous across politicians but quite heterogeneous across topics. On the politician side, regressing two indicators for whether the tweet is blame or merit on demographics yields no systematic differences (Figure A12). In addition, the distribution of within-member changes among those serving both before and after 2016 is positive and shows only moderate variation (Figure A9). On the topic side, causal rhetoric is concentrated in policy domains rather than sociocultural ones (Figure A10), and most of the increase also comes from policy topics (Figure A11).

Taken together, these decompositions show that the rise of causal rhetoric is not driven by mechanical factors, but instead reflects a pervasive within-politician and within-topic shift in communication style – pointing to blame and merit as strategies, rather than types.

3.3 Causality vs. Purely Affective Messaging

As causal rhetoric rises, a natural question is: what does it replace? Given that politicians deploy blame and merit in a competitive setting, a likely candidate is purely affective messaging.

To test this point, Figure 4a plots the share of blame tweets among those expressing negative sentiment, and the share of merit tweets among those expressing positive sentiment, split by presidencies. The pattern is clear: blame increasingly absorbs the negative sentiment space (23% Obama, 42% Biden), while merit does the same for positive sentiment (13% Obama, 31% Biden).

In turn, we ask whether the growing use of causal rhetoric has reshaped how affective polarization is expressed among elites. The evidence suggests it has.

Figure 4b decomposes partisan difference in sentiment between the ruling and the opposing party across presidencies into causal and non-causal tweets. First, consistent with the observation that Congress has grown more polarized over time (DeSilver, 2022), the overall sentiment gap between parties has widened – from roughly 0.1 under Obama to nearly 0.3 under Biden. More strikingly, this increase is entirely driven by causal tweets: by the Biden presidency, sentiment differences in non-causal tweets are indistinguishable from zero, while causal tweets display a divide approaching 0.6. Taken together, these findings suggest that causal rhetoric is not only rising in prominence but increasingly serves as the primary vehicle for elite-level affective polarization.

Appendix B.3 provides different robustness checks and additional analyses for these findings, all supporting the idea that causal rhetoric has become an increasingly important strategic vehicle for elite polarization.

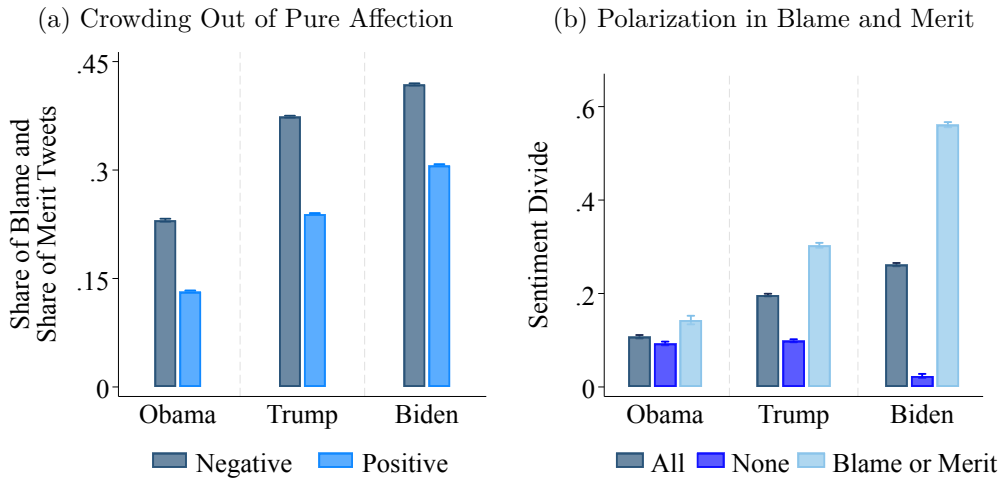
4 Revenues: Causal Effects on Small Donations

The supply of causal rhetoric can be understood through the lens of costs and benefits, as in a standard production problem (Aridor et al., 2024). Following this idea, we causally identify that blame increases aggregate revenues from small donations, while merit has no average effect. A decomposition reveals distinct mechanisms: blame operates at the extensive margin, by raising the number of donors; merit operates at the intensive margin, by increasing the average donation size. Moreover, responses vary across donor groups: extremes respond to blame, moderates to merit. Finally, in line with the rise in the supply of blame and merit, these patterns intensify over time.

4.1 Empirical Strategy

To assess the causal effect of blame and merit on donations, we exploit cross-county variation in Twitter penetration across the United States. The intuition is straightfor-

Figure 4: Affective Messaging



Notes: Panel (a) presents the share of blame tweets among those expressing negative sentiment, and the share of merit tweets among those expressing positive sentiment, separately for each presidency. Panel (b) presents the difference in average standardized sentiment between tweets posted by members of the ruling party and those posted by members of the opposition, separately for each presidency. We compute this difference across three subsamples: all tweets, tweets that are neither blame nor merit, and tweets that are either blame or merit. In both panels, bars represent 95 percent confidence intervals.

ward: in counties where Twitter use is more widespread, residents are more likely to be exposed to the blame and merit messaging supplied by politicians. We then estimate how, for each politician, donations from each county respond to a given level of blame and merit, as a function of local Twitter penetration.

Empirically, we leverage the richness of campaign finance data, which reports donors' locations, to construct a politician-by-county-by-month panel, tracking the donations received by each politician from each county in each month in which they tweet. Then, we estimate the following specification:

$$\begin{aligned} y_{icm} = {} & \beta_1(\text{Blame}_{im} \times \text{Users}_c) + \beta_2(\text{Merit}_{im} \times \text{Users}_c) \\ & + \gamma_1(\text{Blame}_{im} \times \mathbf{X}_c) + \gamma_2(\text{Merit}_{im} \times \mathbf{X}_c) \\ & + \delta_1(\text{Sentiment}_{im} \times \text{Users}_c) + \lambda_{ic} + \mu_{im} + \eta_{mc} + \varepsilon_{icm}. \end{aligned} \tag{3}$$

Here, y_{icm} is our outcome of interest, log+1 donations revenue, log+1 number of donors or log+1 average amount donated per donor – each measured for politician i in county c and month m; Blame_{im} and Merit_{im} measure the share of blame and merit tweets posted by politician i in month m; and Users_c captures the log+1 number of Twitter users in county c; λ_{ic}, μ_{im}, and η_{mc} denote politician-by-county, politician-by-month, and county-by-month fixed effects. In addition, we take two additional precautions in our specification. First, to isolate a Twitter-specific channel from other confounders, we interact the share of blame and merit tweets with a rich set of 28 cross-sectional county-level controls ($\mathbf{X}_c$) borrowed from (Fujiwara et al., 2024) and presented in Table A5.

Second, we include controls for the average sentiment of the politician's tweets in month m, (Sentiment_{im}), ensuring that the estimated effects are not merely capturing variation in positive or negative sentiment. Exploiting this geographical design allows us to include a rich set of fixed effects that can absorb some confounding factors in our analysis: μ_{im} allows us to absorb all shocks to a politician's popularity, λ_{ic} nets out differences in popularity across counties, and finally η_{mc} allows for a county-level shock in donations to all candidates.

A central concern is that Twitter penetration may itself be endogenous to local political or economic conditions. Counties with higher partisan competition or greater donor capacity may have adopted Twitter earlier, and demographics such as income or education may correlate both with penetration and political engagement. If so, our estimates could confound the causal effect of rhetoric with pre-existing differences in demand for political communication. To address these issues, we leverage quasi-exogenous variation in Twitter penetration induced by the platform's early diffusion following the South-by-Southwest (SXSW) festival in 2007, a well-documented shock that catalyzed the platform's early diffusion in the United States (Fujiwara et al., 2024; Müller and Schwarz, 2023). Specifically, we instrument the county-level Twitter users with the number of followers of the official SXSW Twitter account in 2007. This instrument provides plausibly exogenous variation in the intensity of Twitter use across counties, independent of local determinants of political donations.[22]

In practice, we instrument Users_c with the log+1 number of SXSW followers in 2007 in county c ($\text{SXSWFollowers2007}_c$). Then, we consider a standard Two-Stage Least Squares framework, where the interactions involving the level of Twitter penetration of Equation 3 are instrumented with their counterparts based on $\text{SXSWFollowers2007}_c$. In our context, the exclusion restriction amounts to saying that SXSW-induced Twitter adoption affects county-level donations to member i only by scaling exposure to that member's tweets and not through any other channel that directly shifts giving. Following Fujiwara et al. (2024), we also include the interaction of our main regressors of interest with pre-2007 followers of the SXSW account. This ensures that the identifying variation comes from counties that are also similar in observable characteristics. Furthermore, including these interacted controls in the regression can lend credibility to our identifying assumption. Suppose that counties with an interest in SXSW's Twitter account during the early years of the platform also differ in (unobservable) characteristics that predict returns to blame and merit. Then, the coefficients of the pre-2007 regressors should be similar to the main ones. However, we mainly estimate small and non-significant coefficients for these placebo checks.

[22]This instrument is widely used in the literature (Boken et al., 2023, e.g.,). We refer to the original papers for extensive evidence of both its relevance for the growth of local Twitter activity and of its plausible exogeneity.

4.2 Total Revenues, Mobilization and Fidelization

We begin by estimating Equation 3 using total donation revenue as the outcome variable.[23] As shown in Figure 5a, a one standard deviation increase in the monthly share of blame tweets raises donations by about 0.64 percent for each log-point increase in local Twitter penetration. The effect is economically meaningful: in the average county, this translates into a 3.4 percent increase in contribution revenue. By contrast, merit-oriented rhetoric has no statistically or economically significant effect on overall donation revenues.

Having established the aggregate effect, we next ask how these revenues accrue. Intuitively, donor fundraising can expand along two margins: the extensive margin, through the number of unique donors donating to the politician (mobilization), and the intensive margin, through the average amount donated by each of these donors (fidelization). To this end, we replicate the main specification using the number of unique donors and the average total amount donated per unique donor as separate outcomes.

Results are shown in Figures 5b and 5c. Blame has a strong and statistically significant effect on the number of donors: a one standard deviation increase raises the number of unique donors by about 0.42 percent per log-point of penetration. In the average county, this translates into an increase in the number of donors by 2.2 percent. Merit, on the contrary, has a small and non-significant negative effect on this margin. The pattern reverses for the average total amount donated per donor. Blame has a positive but extremely noisy impact, while merit increases the average donation by about 0.16 percent per log-point of penetration, which in the average county translates into an increase of 0.8 percent.

To clarify the magnitude of our estimates, we calculate the persuasion rate (DellaVigna and Gentzkow, 2010), focusing on the results related to the number of donors.[24] We find a persuasion rate of 1.3%,[25] which is slightly higher than the one associated with political advertising (Spenkuch and Toniatti, 2018) and opening a Twitter account (Petrova et al., 2021), but still lower than the effect of virality (Boken et al., 2023) and the average rates reported in DellaVigna and Gentzkow (2010).

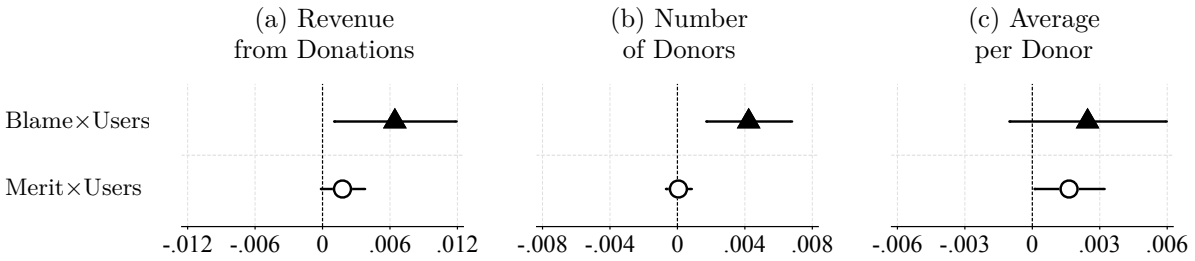
Taken together, these results suggest that, while blame is more effective than merit in the aggregate, the two rhetorical styles serve complementary purposes: blame helps politicians reach more donors, while merit helps deepen engagement with existing supporters. We conclude the Section by providing evidence in support of the mobilization

[23] Table A6 presents detailed results. Besides, Table A7 presents the OLS results, Table A8 presents the reduced form results, while Table A9 presents the first stage results. In general, we find coefficients of comparable magnitude between the OLS and the 2SLS results.

[24] As discussed in (Boken et al., 2023), persuasion rates are more conceptually appropriate for decisions represented as binary outcomes.

[25] Appendix B.2 reports the details of how we compute the persuasion rate, detailing how the estimate reported here is a lower bound.

Figure 5: Returns



Notes: The figure presents the estimates of Equation 3, considering as outcome the log-donations revenue, log-donations number, and log-average donation in Panels (a), (b), and (c), respectively. Bars represent 95 percent confidence intervals computed with standard errors clustered at the county level.

mechanism, analyzing heterogeneity in donor responses, and linking the return patterns to the supply dynamics documented above.

4.3 Blame and Virality

If reach is the channel through which blame mobilizes Twitter users and increases the number of donors, we should expect blame tweets to spread more easily than others. To this end, we investigate the relationship between rhetorical style and retweets. Appendix B.4 describes in more detail the data and specifications used.

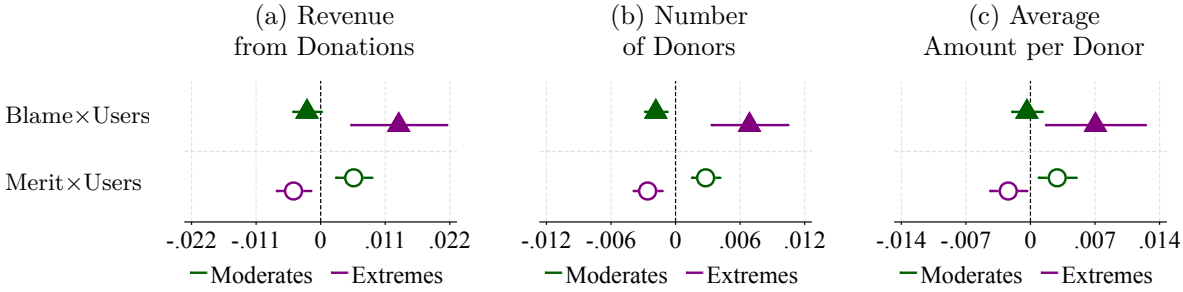
First, relative to non-causal tweets, blame tweets receive about 0.2 standard deviations more retweets, while controlling for the sentiment of the tweet and including politician and topic fixed effects. Second, the effect is concentrated in the upper tail of the popularity distribution: blame is negatively associated with being in the lower deciles of retweets, but increasingly positive in the upper ones – especially in the top decile. Third, a back-of-the-envelope calculation highlights the magnitude: although blame accounts for just 15 percent of tweets, it generates nearly 40 percent of all retweets. Merit shows no comparable pattern. Finally, blame is increasingly associated with engagement and virality over time, in line with its increase in supply by politicians and returns in terms of donations.

Together, these results show that blame is not only more engaging on average, but also disproportionately more likely to go viral. This pattern supports the mobilization channel and aligns our evidence with Boken et al. (2023), who find that small donations spike when a politicians tweet "goes viral."

4.4 Donor-Level Heterogeneity

An additional reason why blame and merit serve complementary purposes (beyond the mobilization vs. fidelization distinction outlined above) is provided by donor-level heterogeneity.

Figure 6: Returns across Donors



Notes: The figure presents the estimates of 3, considering as outcome the log-donations revenue, log-donations number, and log-average donation in Panels (a), (b), and (c), respectively. Each outcome is computed separately for moderate and extreme donors. Bars represent 95 percent confidence intervals computed with standard errors clustered at the county level.

In particular, we draw on an established measure of donors' ideology constructed from campaign finance records and analogous to Nominate scores for legislators (Bonica, 2014). We classify donors as moderates or extremes based on whether the absolute value of their ideological score falls below or above the median within each electoral cycle. We then re-estimate Equation 3, disaggregating donations by donor type. Figure 6 reveals a stark divide: moderates respond strongly positively to merit and are indifferent to blame, while extremes respond markedly positively to blame and negatively to merit.[26]

This heterogeneity suggests that rhetorical strategies are not uniformly effective across the donor base. Instead, blame and merit appeal to ideologically distinct constituencies – revealing, once again, their complementary functions.

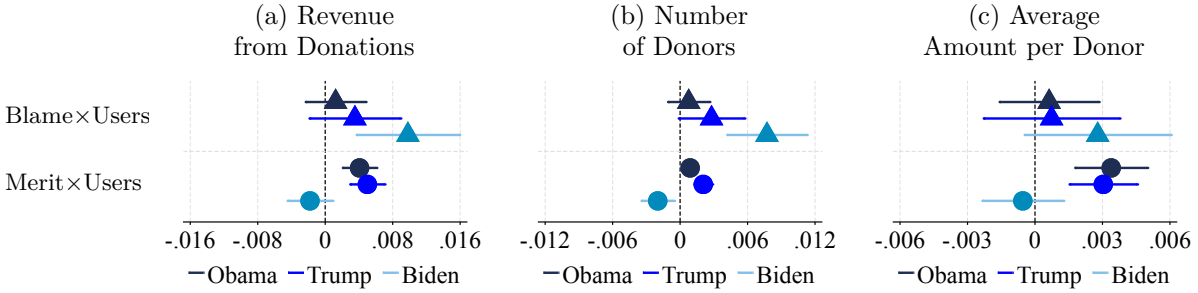
4.5 Linking Supply and Returns

We conclude by examining whether the supply of causal rhetoric over time is reflected in its evolving returns. While this analysis is descriptive and does not aim to establish causal identification, it offers suggestive insights into the dynamics of rhetorical effectiveness. To explore this, we re-estimate Equation 3, allowing the coefficients on the key interaction terms – namely ($\text{Blame}_{im} \times \text{Users}_c$) and ($\text{Merit}_{im} \times \text{Users}_c$) – as well as the corresponding pre-SXSW control interactions, to vary across presidencies. Specifically, we interact these terms with indicator variables for the presidential administration. This specification allows us to examine how the estimated returns to causal rhetoric evolve over time.

The results are presented graphically in Figure 7.[27] Returns to blame increase beginning with Trumps term, while the returns to merit do not display a comparable pattern, starting positive in the first two terms before becoming negative. The increase in returns

[26] Tables A10 and A11 presents detailed results.
[27] Tables A12 to A14 presents detailed results.

Figure 7: Returns over Time



Notes: The figure presents the estimates of Equation 3, considering as outcome the log-donations revenue, log-donations number, and log-average donation in Panels (a), (b), and (c), respectively. The specifications include interactions that allow the effects of blame and merit to vary across the three presidencies. Bars represent 95 percent confidence intervals computed with standard errors clustered at the county level.

to blame is driven primarily by the extensive margin: blame begins to significantly raise the number of donors only during and after the Trump presidency.

This pattern mirrors the supply dynamics documented above and suggests that the payoff to causal rhetoric – particularly blame – rises as it becomes a more entrenched feature of political communication.

5 Costs: Constraints to Credibility

Having established the returns to causal rhetoric, we next turn to its strategic costs. Once we condition on power status, a systematic trade-off emerges between blame and merit: opposition members deploy blame, while incumbents emphasize merit. We interpret this as a reputational cost. When in government, credibly blaming others becomes more difficult, whereas taking merit is easier; and vice versa. Consistent with this interpretation, we show that rhetorical choices respond to legislative activity, as blame rises when attacking bills sponsored by the opposing party, while merit rises when defending one's own.

5.1 The Trade-Off between Blame and Merit

Figure 8 plots the share of merit versus blame tweets for each member of Congress, color-coded by power status and separated by presidency. A clear contrast emerges between the Obama years – prior to the widespread supply of causal rhetoric – and the Trump and Biden years that follow. In the earlier period, no systematic relationship is visible.[28] In the later period, a clear substitution pattern emerges, revealing a trade-off between blame and merit.

[28]If anything, the slope is weakly positive, reflecting differences between high-usage and low-usage members – those who employ both merit and blame versus those who use neither.

This trade-off is striking, given that both blame and merit are associated with distinct political returns – blame with total revenue and mobilization, merit with fidelization, extremes responding to one, moderates to the other. Why, then, do politicians not use both? The pattern suggests a cost.

To unpack the nature of this cost, we quantify the determinants of the blame-merit trade-off by studying the correlation of blame and merit usage with politicians' characteristics in Figure A12. While demographics have little explanatory power, ideological extremism is the strongest predictor of blame and merit, echoing the donor-side heterogeneity in Section 4.4. But power status – a dynamic attribute – is the dominant driver: opposition members consistently show a wider gap between blame and merit, and this pattern strengthens over time (mirroring the clustering in Figure 8).

Importantly, the blame-merit trade-off is not equivalent to a simple shift between negative and positive sentiment. In Figure A14, we show that the share of blame tweets among those expressing negative sentiment increases when politicians are in opposition compared to when they are in power. A symmetric pattern holds for merit: among tweets expressing positive sentiment, the share of merit increases when politicians are in power relative to when they are in opposition. This pattern implies that causal rhetoric is more responsive to power status than sentiment itself. When government changes hands, politicians adjust their use of blame and merit more than they adjust the underlying sentiment of their communication.

Finally, rhetorical adjustments are swift: event-study estimates (Figure A13) show that members quickly pivot between blame and merit as control of government changes hands.

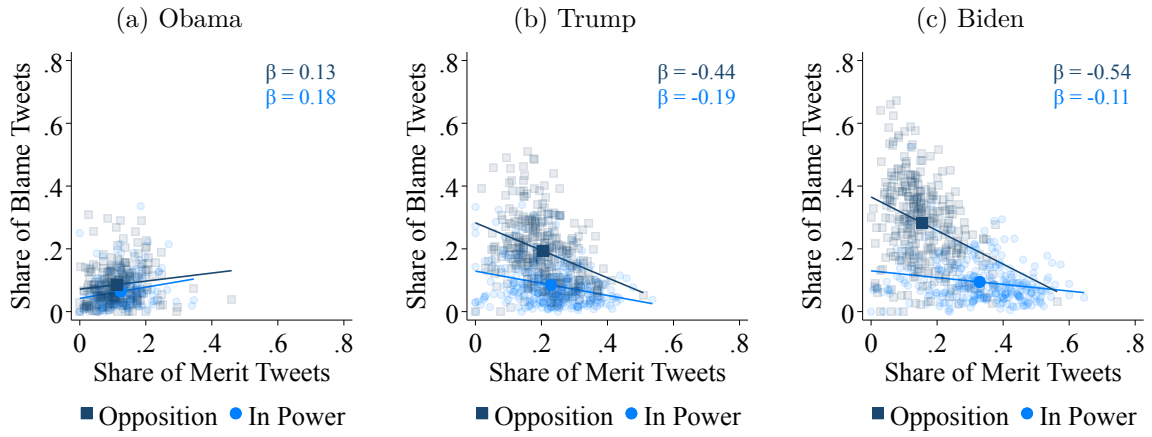
These patterns suggest that the cost appears to arise from a credibility constraint. Politicians cannot freely assign responsibility when they hold legislative power: if they are the ones in power, they cannot credibly blame; if they are not, they cannot credibly claim merit. In this sense, rhetorical choices reflect not just incentives, but strategic limits – what can be said credibly, given one's institutional position.

5.2 Rhetorical Style and Legislative Activity

To explore this mechanism more directly, we turn to a more fine-grained measure of political activity than general power status. We use bill proposals as a proxy for observable political action. Specifically, we estimate the following specification:

$$\begin{aligned} y_{ipt} = {} & \beta_1 \text{Opposing}_{pt} \times D_t^{\text{Obama}} + \beta_2 \text{ Opposing}_{pt} \times D_t^{\text{Trump}} + \beta_3 \text{ Opposing}_{pt} \times D_t^{\text{Biden}} \\ & + \gamma_1 \text{Own}_{pt} \times D_t^{\text{Obama}} + \gamma_2 \text{ Own}_{pt} \times D_t^{\text{Trump}} + \gamma_3 \text{ Own}_{pt} \times D_t^{\text{Biden}} \\ & + \lambda_p + \mu_{week} + \varepsilon_{ipt}. \end{aligned} \tag{4}$$

Figure 8: Trade-Off



Notes: Each panel presents a scatterplot at the politician level showing the share of tweets classified as merit and blame during the respective presidency. The solid line is from a linear regression of the share of blame tweets over the share of merit tweets for politicians in power (opposition), with the estimate reported in the top right corner. In all panels, politicians who posted fewer than 10 tweets during the presidency are excluded, which are 34, 36, and 5 for the Obama, Trump, and Biden presidencies, respectively.

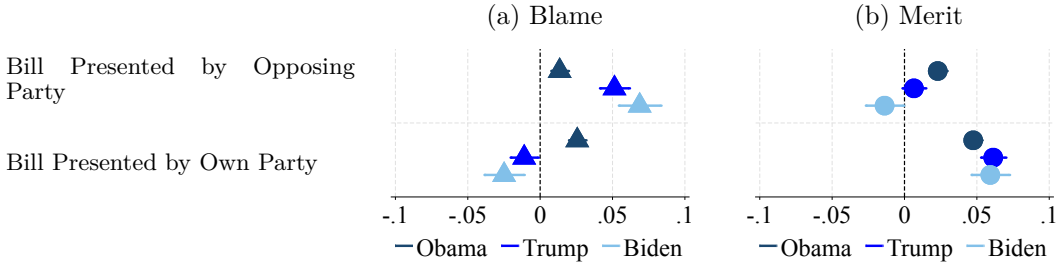
Here y_{ipt} is a binary indicator denoting whether tweet i posted by politician p on date t is blame (or merit); Opposing_{pt} is a binary indicator taking value 1 if a member of the politician's opposing party introduced a bill on day t and Own_{pt} is defined analogously for bills introduced by members of the own party; D_t^{Obama}, D_t^{Trump} and D_t^{Biden} are indicators for whether the sitting president on day t is Obama, Trump or Biden, respectively; λ_p, and μ_{week} are politician and week fixed effects, respectively. This design allows us to examine whether politicians' use of blame and merit systematically relates to legislative activity by their own or the opposing party – and how this relationship varies across presidencies. The inclusion of fixed effects captures time-invariant differences across politicians and absorbs common shocks in the news or legislative agenda at the week level.

Figure 9 summarizes our findings.[29] Panel (a) shows that, before 2017, legislative activity was not associated with an increase in the supply of blame. Starting from 2017, however, a clear pattern emerges: when the opposing party introduces a bill, the probability of posting a blame tweet rises to about 4 percentage points under Trump and to about 6 percentage points under Biden. The effect reverses when the bill is introduced by ones own party. Merit shows a similar pattern, with one distinction: already under Obama, merit was about 3 percentage points more likely when a bill came from the politicians own party, and this effect strengthened to 4 percentage points in subsequent terms.

A natural concern is that our bills results could be driven by shifts in overall tone, with politicians tweeting more negatively when the other party introduces bills and more positively when their own party does, rather than by changes in causal attribution. To

[29] Table A15 presents detailed results.

Figure 9: Legislative Activity



Notes: The figure presents the estimates of Equation 4, considering as outcomes a binary indicator for whether the tweet is blame and a binary indicator for whether the tweet is merit in Panels (a) and (b), respectively. Bars represent 95 percent confidence intervals computed with standard errors clustered at the day level.

show the distinct margin of causal rhetoric, we address this concern by limiting our analysis to only negative (positive) tweets when looking at changes in blame (merit). In practice, we estimate Equation 4 only on tweets classified as negative (positive) when the outcome is a binary indicator denoting whether the tweet is blame (merit) or not. We report the results in Figure A15 and Table A16. They are in line with the discussion above. Within negative tweets, the share of blame increases when the opposing party presents bills, and it does so increasingly over time, while the opposite holds when the bills are presented by their own party. The patterns are similar for merit. This shows that politicians strategically respond more within the causal dimension than the non-causal one.

Together, these results show that causal rhetoric follows a credibility logic – claiming merit requires ownership of policy, while assigning blame is only available as a tool of opposition.

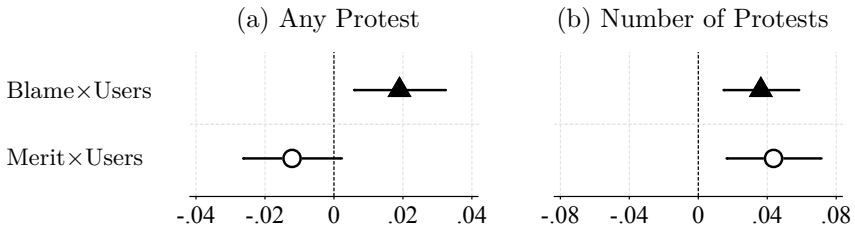
6 Societal Outcomes: Protests and Polarization

Having analyzed the supply, returns, and costs of causal rhetoric, we now ask whether it produces externalities on societal outcomes. First, using our geographic design, we show that blame increases the incidence of protests at the county level, while merit raises the number of protests – resonating with the mobilization and fidelization channels. Second, we find that exposure to blame is associated with greater affective polarization, lower trust in government, and lower perceived government effectiveness.

6.1 Protests

A natural political outcome to examine in our context is protest activity, which reflects direct political engagement rather than financial support, offering a chief example of offline political behavior.

Figure 10: Protests



Notes: The figure presents the estimates of β_1 and β_2 from Equation 5, considering as outcome a binary indicator for whether any protest occurred and the log number of protests Panels (a), and (b) respectively. Bars represent 95 percent confidence intervals computed with standard errors clustered at the county level.

To test the effects of causal rhetoric, we estimate a version of our baseline specification using county-level protest data as the outcome.

$$\begin{aligned} y_{cm} = \; & \beta_1(\text{Blame}_m \times \text{Users}_c) + \beta_2(\text{Merit}_m \times \text{Users}_c) \\ & + \gamma_1(\text{Blame}_m \times \mathbf{X}_c) + \gamma_2(\text{Merit}_m \times \mathbf{X}_c) \\ & + \delta_1(\text{Sentiment}_m \times \text{Users}_c) + \lambda_c + \mu_m + \varepsilon_{cm}. \end{aligned} \tag{5}$$

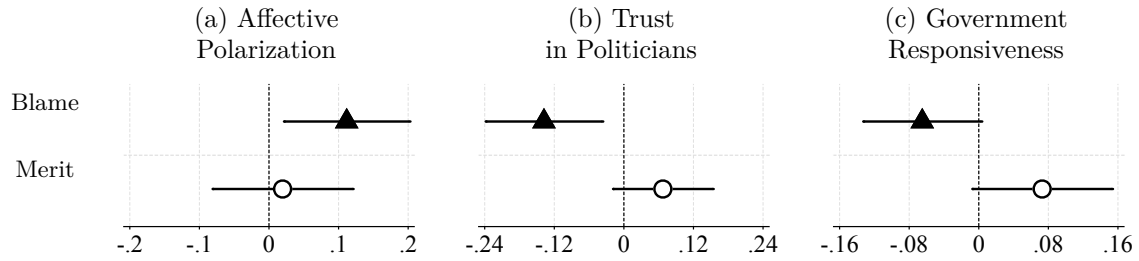
Here, y_{cm} is our outcome of interest, namely, either whether a protest has taken place in month m and county c or the log+1 of the number of protests occurring in month m and county c; Blame_m and Merit_m measure the overall share of blame and merit tweets posted by politicians in month m; otherwise, the specification follows the one introduced in Section 4.

We report results in Figure 10.[30] A one standard deviation increase in the share of blame tweets raises the likelihood of a protest by 1.9 percentage points per log-point of Twitter penetration. This effect is sizable: in the average county, this translates into a 10.1 percentage point increase in protest probability. By contrast, merit has a smaller, negative, and statistically insignificant effect. This finding mirrors the mobilization effect documented in donation behavior: blame increases low-cost participation. To capture the intensive margin, we examine the number of protests. Here, consistent with the fidelization effect, we find that merit plays a role. A one standard deviation increase in merit is associated with a 4.4 percent rise in the number of protests per log-point of Twitter penetration, which, in the average county, implies an increase in the number of protests by 22.6 percent. In this case, the effect of blame is comparable.

To put the magnitudes estimated above in context, we relate them to previous results in the literature. Enikolopov et al. (2020) find that a 10% increase in social media penetration in Russia increases the probability of a protest by 4.6% and the number of protestors by 19%, while Gylfason (2023) finds that a 1% increase in Twitter penetration increases the frequency of protests by 1% in the United States. Overall, these results

[30] Table A17 presents detailed results.

Figure 11: Attitudes



Notes: The figure presents the estimates of Equation 6, considering as outcome affective polarization, trust in politicians, and perceived government responsiveness in Panels (a), (b), and (c), respectively. Bars represent 95 percent confidence intervals computed with standard errors clustered at the state level.

show that the penetration effect on protests is sizeable, consistently with our estimates.

6.2 Affective Polarization and Attitudes Toward Government

A second natural hypothesis is that politicians' rhetorical style shapes voters' attitudes toward peers and government institutions.

To examine this possibility, we analyze data from America's Political Pulse (Westwood et al., 2024), which consists of a weekly survey administered to a nationally representative sample of U.S. voters since late 2022. In particular, we focus on answers regarding affective polarization, trust in politicians, and perceived government responsiveness.[31] Then, we correlate respondents' beliefs and attitudes with the share of merit and blame tweets posted by politicians in their state. Essentially, we estimate the following specification:

$$y_{sw} = \beta_1 \text{Blame}_{sw} + \beta_2 \text{Merit}_{sw} + \lambda_w + \varepsilon_{sw}. \tag{6}$$

Here y_{sw} is the average level of our attitude of interest – namely, affective polarization, trust in politicians, and perceived government responsiveness – in state s and week w; Blame_{sw} and Merit_{sw} are the share of blame and merit tweets posted by politicians elected in state s during week w; and λ_w are week fixed effects. All the outcomes and regressors are standardized to make them comparable.

Figure 11 presents the results.[32] First, we find that the amount of blame is positively associated with greater affective polarization. In addition, both merit and blame correlate with downstream political attitudes in expected ways: blame is linked to lower political trust and reduced perceptions of government responsiveness, while merit is associated with more positive views.

[31] Details about the survey questions used are provided in Table A21

[32] Table A22 presents detailed results.

7 Conclusion

There are at least two ways a politician can try to shape public opinion. They can appeal to emotion alone, or they can weave sentiment into a causal explanation – telling voters who deserves blame or credit for the current state of affairs. While emotional and affective messaging has received sustained attention, we know less about how politicians use causal claims.

This paper took a first step in that direction. We showed that causal rhetoric rose sharply over the past decade, displacing affective messaging and becoming a central medium for elite polarization. We studied its benefits and costs: blame increases donation revenues by expanding donor count and spreading virally, while merit raises the average donation size. Rhetorical choices, however, are constrained by credibility – shaped in part by whether a politician supports or opposes the policy in question. Finally, we showed that causal rhetoric influences real-world outcomes, including protest activity, affective polarization, and institutional trust.

In future work, we plan to explore the socio-psychological foundations of how individuals respond to causal rhetoric, blame, and merit, as well as their economic consequences. Preferences for causal rhetoric may reflect deeper traits – for instance, a stronger consequentialist rather than deontological outlook (Awad et al., 2020; Graham et al., 2009; Piazza and Sousa, 2014). Similarly, heightened responsiveness to blame may reflect underlying negativity bias in some constituencies (Baumeister et al., 2001; Hibbing et al., 2014; Rozin and Royzman, 2001). These connections warrant further study, as recent economic research has begun to engage with related questions (Bénabou et al., 2024).

Bibliography

Alesina, A., A. Miano, and S. Stantcheva (2023). Immigration and redistribution. *The Review of Economic Studies 90*(1), 1–39.

Algan, Y., E. Davoine, T. Renault, and S. Stantcheva (2025). Emotions and policy views. Technical report, Mimeo.

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020). The welfare effects of social media. *American economic review 110*(3), 629–676.

Andre, P., I. Haaland, C. Roth, and J. Wohlfart (2021). Narratives about the macroeconomy.

Aridor, G., R. Jiménez-Durán, R. Levy, and L. Song (2024). The economics of social media. *Journal of Economic Literature 62*(4), 1422–1474.

Ash, E. and S. Hansen (2023). Text algorithms in economics. *Annual Review of Economics 15*(1), 659–688.

Awad, E., S. Dsouza, A. Shariff, I. Rahwan, and J.-F. Bonnefon (2020). Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proceedings of the National Academy of Sciences 117*(5), 2332–2337.

Barberá, P., A. Casas, J. Nagler, P. J. Egan, R. Bonneau, J. T. Jost, and J. A. Tucker (2019). Who leads? who follows? measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review 113*(4), 883–901.

Barron, K. and T. Fries (2024a). Narrative persuasion. Technical report, WZB Discussion Paper.

Barron, K. and T. Fries (2024b). Narrative persuasion: A brief introduction encyclopedia of experimental social science.

Baumeister, R. F., E. Bratslavsky, C. Finkenauer, and K. D. Vohs (2001). Bad is stronger than good. *Review of general psychology 5*(4), 323–370.

Bellodi, L., M. Morelli, A. Nicolo, and P. Roberti (2023). The shift to commitment politics and populism: Theory and evidence. *BAFFI CAREFIN Centre Research Paper* (204).

Bénabou, R., A. Falk, and L. Henkel (2024). Ends versus means: Kantians, utilitarians, and moral decisions. Technical report, National Bureau of Economic Research.

Bilotta, F. and G. Manferdini (2024). Coarse memory and plausible narratives. *Available at SSRN 4700043*.

Boken, J., M. Draca, N. Mastrorocco, and A. Ornaghi (2023). The returns to viral media: the case of us campaign contributions.

Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science 58*(2), 367–386.

Bonica, A. (2024). Database on ideology, money in politics, and elections: Public version 4.0. [Computer file].

Bouton, L., J. Cagé, E. Dewitte, and V. Pons (2022). Small campaign donors. Technical report, National Bureau of Economic Research.

Boyer, P., G. Gauthier, Y. L. Yaouanq, V. Rollet, and B. Schmutz (2024). The lifecycle of protests in the digital age.

Burnham, M., K. Kahn, R. Y. Wang, and R. X. Peng (2024). Political debate: Efficient zero-shot and few-shot classifiers for political text. *arXiv preprint arXiv:2409.02078*.

Bursztyn, L., A. Rao, C. Roth, and D. Yanagizawa-Drott (2023). Opinions as facts. *The Review of Economic Studies 90*(4), 1832–1864.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal 17*(2), 372–404.

Camacho-Collados, J., K. Rezaee, T. Riahi, A. Ushio, D. Loureiro, D. Antypas, J. Boisson, L. Espinosa-Anke, F. Liu, E. Martínez-Cámara, et al. (2022). Tweetnlp: Cutting-edge natural language processing for social media. *arXiv preprint arXiv:2206.14774*.

Campante, F., R. Durante, and A. Tesei (2022). Media and social capital. *Annual review of economics 14*(1), 69–91.

Chater, N. and G. Loewenstein (2016). The under-appreciated drive for sense-making. *Journal of Economic Behavior & Organization 126*, 137–154.

CongressTweets (2023). https://github.com/alexlitel/congresstweets.

Cormack, L. (2025). DCinbox: Official e-newsletters from every member of congress. https://www.dcinbox.com/about/.

Crowd Counting Consortium (2025). Crowd counting consortium. https://ash.harvard.edu/programs/crowd-counting-consortium/. Harvard Kennedy School and University of Connecticu.

Davidson, T., D. Warmsley, M. Macy, and I. Weber (2017). Automated hate speech detection and the problem of offensive language. In *Proceedings of the international AAAI conference on web and social media*, Volume 11, pp. 512–515.

DellaVigna, S. and M. Gentzkow (2010). Persuasion: empirical evidence. *Annu. Rev. Econ. 2*(1), 643–669.

DellaVigna, S. and E. Kaplan (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics 122*(3), 1187–1234.

Demszky, D., N. Garg, R. Voigt, J. Zou, M. Gentzkow, J. Shapiro, and D. Jurafsky (2019). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

DeSilver, D. (2022). *The polarization in todays Congress has roots that go back decades*. Pew Research Center.

Di Tella, R., R. Kotti, C. Le Pennec, and V. Pons (2023). Keep your enemies closer: strategic platform adjustments during us and french elections. Technical report, National Bureau of Economic Research Cambridge, MA.

Eliaz, K. and A. Rubinstein (2025). Wasonian persuasion.

Eliaz, K., R. Spiegler, and S. Galperti (2023). False Narratives and Political Mobilization. CEPR Discussion Papers 17832, C.E.P.R. Discussion Papers.

Enikolopov, R., A. Makarin, and M. Petrova (2020). Social media and protest participation: Evidence from russia. *Econometrica 88*(4), 1479–1514.

Enikolopov, R., M. Petrova, and E. Zhuravskaya (2011). Media and political persuasion: Evidence from russia. *American economic review 101*(7), 3253–3285.

Enke, B. (2020). Moral values and voting. *Journal of Political Economy 128*(10), 3679–3729.

Fujiwara, T., K. Müller, and C. Schwarz (2024). The effect of social media on elections: Evidence from the united states. *Journal of the European Economic Association 22*(3), 1495–1539.

Gehring, K. and M. Grigoletto (2023). Analyzing climate change policy narratives with the character-role narrative framework. Technical report, CESifo Working Paper.

Gennaro, G. and E. Ash (2022). Emotion and reason in political language. *The Economic Journal 132*(643), 1037–1059.

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–574.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from us daily newspapers. *Econometrica 78*(1), 35–71.

Goetzmann, W. N., D. Kim, and R. J. Shiller (2022). Crash narratives. Technical report, National Bureau of Economic Research.

Graeber, T., C. Roth, and C. Schesch (2024). Explanations. Technical report, CESifo Working Paper.

Graeber, T., F. Zimmermann, and C. Roth (2022). Stories, statistics, and memory.

Graham, J., J. Haidt, and B. A. Nosek (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology 96*(5), 1029.

Gylfason, G. (2023). From tweets to the streets: Twitter and extremist protests in the united states. *Available at SSRN 4551807*.

Halberstam, Y. and B. Knight (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of public economics 143*, 73–88.

He, H. and E. A. Garcia (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering 21*(9), 1263–1284.

Hibbing, J. R., K. B. Smith, and J. R. Alford (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and brain sciences 37*(3), 297–307.

Hüning, H., L. Mechtenberg, and S. Wang (2022). Using arguments to persuade: Experimental evidence. *Available at SSRN 4244989*.

Hutto, C. and E. Gilbert (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Volume 8, pp. 216–225.

Kendall, C. W. and C. Charles (2023). Causal narratives. Technical report, National Bureau of Economic Research.

Lau, R. R. and I. B. Rovner (2009). Negative campaigning. *Annual review of political science 12*(1), 285–306.

Lazarus, R. S. (1991). Cognition and motivation in emotion. *American psychologist 46*(4), 352.

Lombrozo, T. and N. Vasilyeva (2017). Causal explanation. *The Oxford handbook of causal reasoning*, 415.

Loureiro, D., F. Barbieri, L. Neves, L. E. Anke, and J. Camacho-Collados (2022). Timelms: Diachronic language models from twitter. *arXiv preprint arXiv:2202.03829*.

Macaulay, A. and W. Song (2023). Narrative-driven fluctuations in sentiment: Evidence linking traditional and social media. Technical report, Bank of Canada Ottawa.

Malle, B. F., S. Guglielmo, and A. E. Monroe (2014). A theory of blame. *Psychological Inquiry 25*(2), 147–186.

Manski, C. F. (1995). *Identification problems in the social sciences*. Harvard University Press.

Müller, K. and C. Schwarz (2023). From hashtag to hate crime: Twitter and antiminority sentiment. *American Economic Journal: Applied Economics 15*(3), 270–312.

Nguyen, D. Q., T. Vu, and A. T. Nguyen (2020). Bertweet: A pre-trained language model for english tweets. *arXiv preprint arXiv:2005.10200*.

OpenSecrets (2025). Expenditures 2024 cycle. https://www.opensecrets.org/campaign-expenditures?cycle=2024.

Pennebaker, J. W., R. L. Boyd, K. Jordan, and M. Blackburn (2022). Liwc 22. https://www.liwc.app/. Linguistic Inquiry and Word Count 22. Accessed: 2025-03-22.

Petrova, M., A. Sen, and P. Yildirim (2021). Social media and political contributions: The impact of new technology on political competition. *Management Science 67*(5), 2997–3021.

Piazza, J. and P. Sousa (2014). Religiosity, political orientation, and consequentialist moral thinking. *Social Psychological and Personality Science 5*(3), 334–342.

Reimers, N. and I. Gurevych (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Rotesi, T. (2019). The impact of twitter on political participation. *Bocconi University, Milan, Italy Working paper*.

Rozin, P. and E. B. Royzman (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review 5*(4), 296–320.

Sloman, S. A. and D. Lagnado (2015). Causality in thought. *Annual review of psychology 66*, 223–247.

Spenkuch, J. L. and D. Toniatti (2018). Political advertising and election results. *The Quarterly Journal of Economics 133*(4), 1981–2036.

Talat, Z. and D. Hovy (2016). Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In *Proceedings of the NAACL student research workshop*, pp. 88–93.

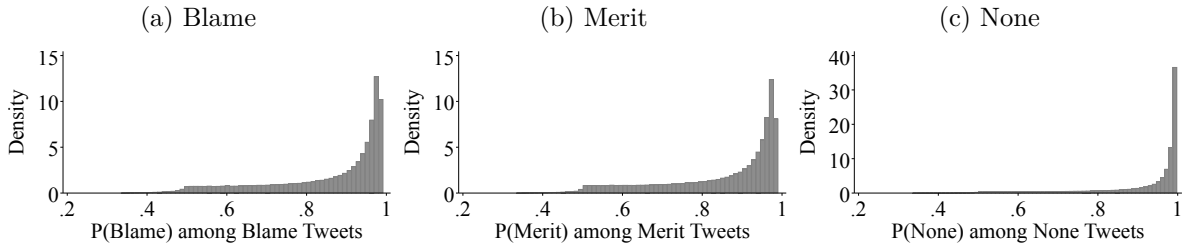
Westwood, S., Y. Lelkes, and M. Wetzel (2024). America's political pulse: Elected officials.

Wojcieszak, M., A. Casas, X. Yu, J. Nagler, and J. A. Tucker (2022). Most users do not follow political elites on twitter; those who do show overwhelming preferences for ideological congruity. *Science advances 8*(39), eabn9418.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual review of economics 12*(1), 415–438.

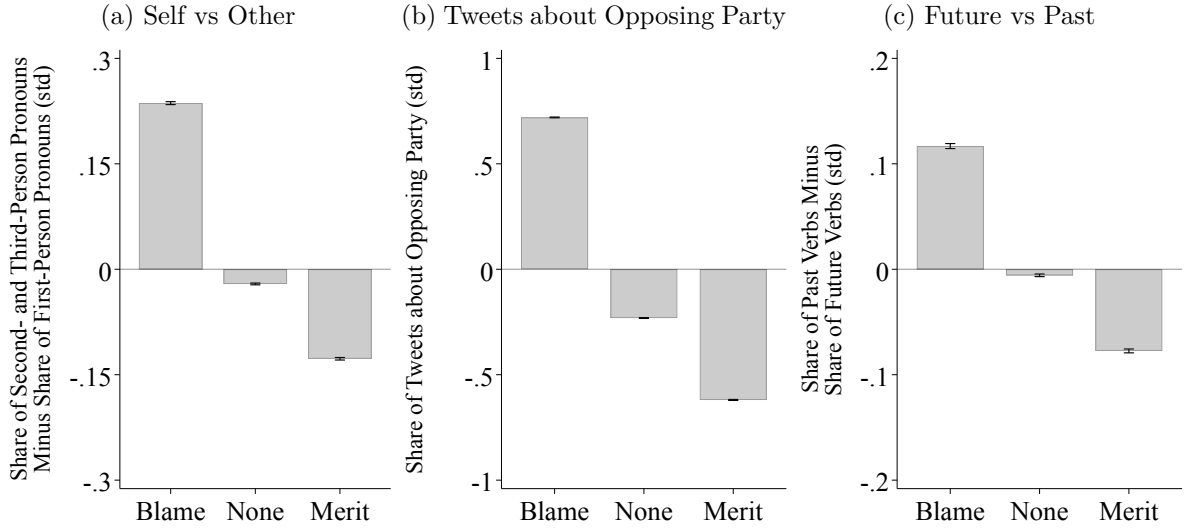
A Additional Tables and Figures

Figure A1: Distributions



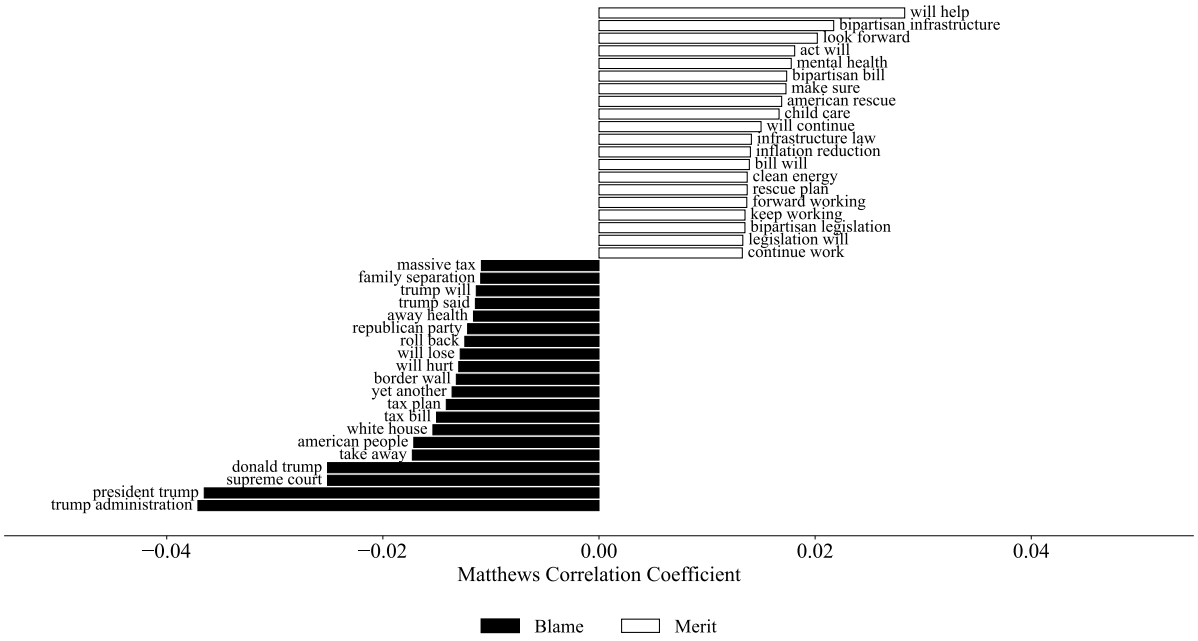
Notes: Panel (a) presents the distribution of the probability that our classifier assigns to blame among tweets classified as blame, that is, among those tweets for which $P(\text{Blame}) > \max\{P(\text{Merit}), P(\text{None})\}$. Panel (b) does the same thing for merit, and Panel (c) does the same for none.

Figure A2: Linguistic Features



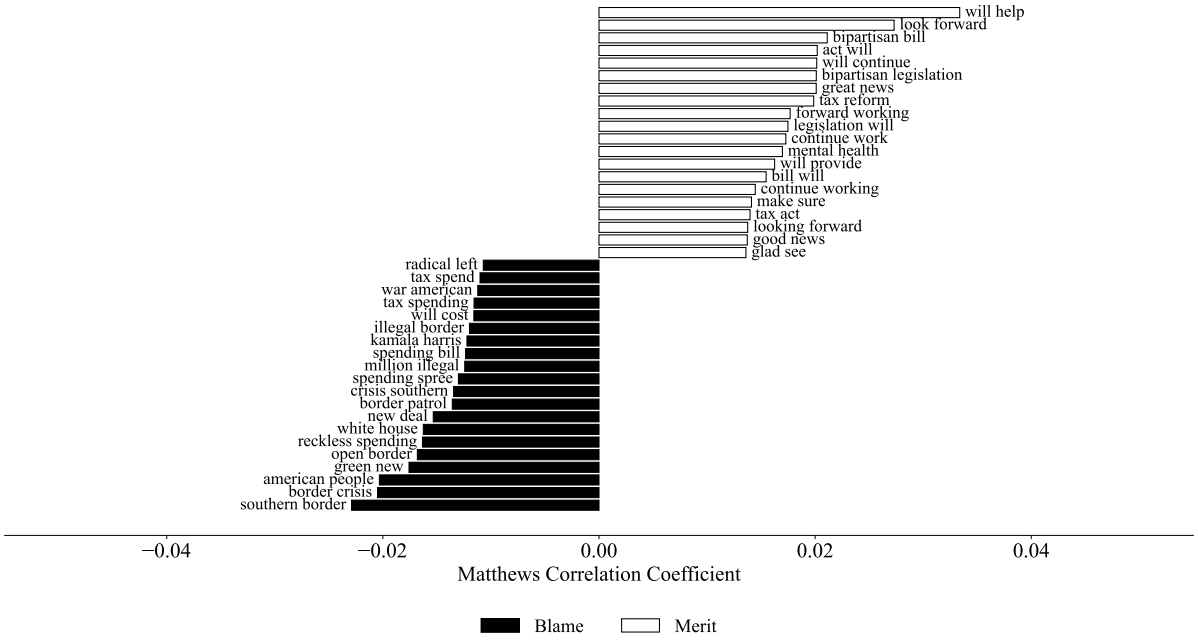
Notes: The figure shows the distribution, across rhetorical styles, of the variable described on the y-axis. Bars represent 95% confidence intervals.

Figure A3: Merit and Blame Bigrams: Democrats



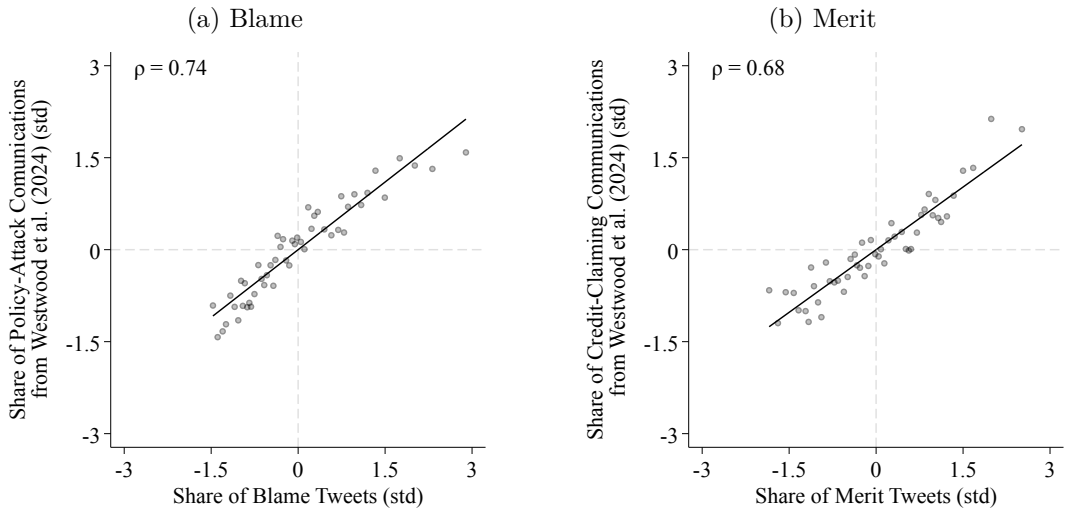
Notes: The figure presents 20 most distinctive bigrams of merit and blame tweets among Democrats, according to their Matthews Correlation Coefficient.

Figure A4: Merit and Blame Bigrams: Republicans



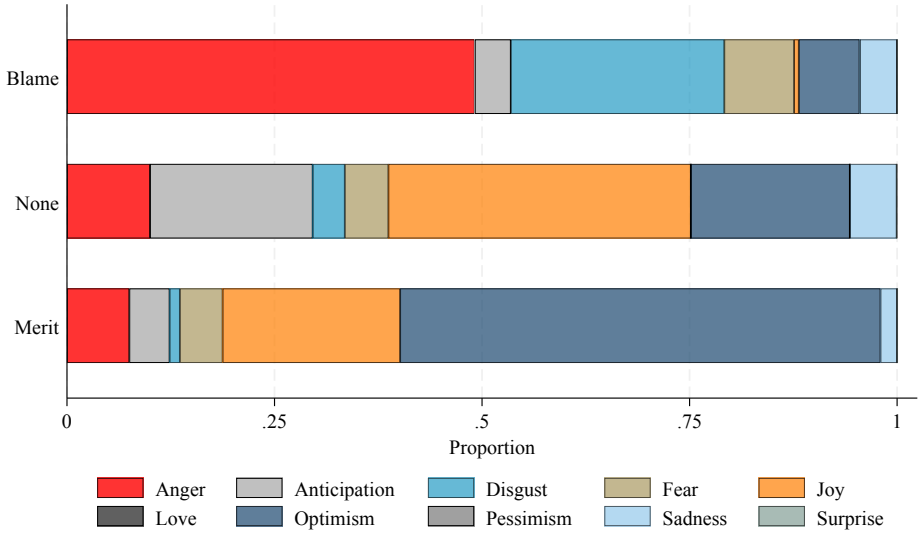
Notes: The figure presents 20 most distinctive bigrams of merit and blame tweets among Republicans, according to their Matthews Correlation Coefficient.

Figure A5: Validation



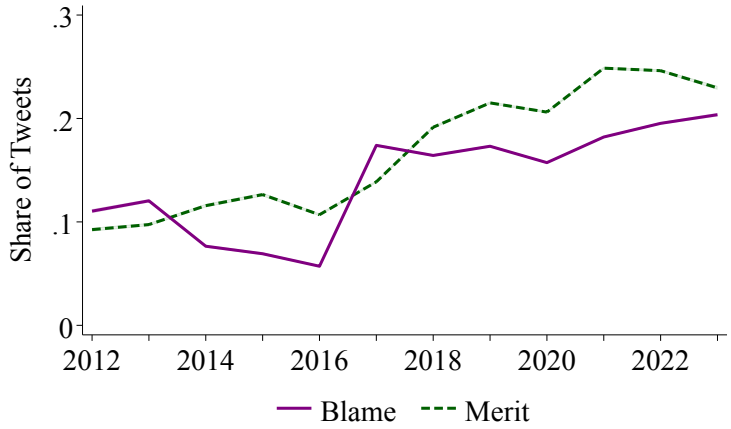
Notes: On the x-axis, we report the share of tweets tagged as blame (or merit) for each politician. On the y-axis, we report the share of policy attack (or credit claiming) communications from Westwood et al. (2024) for each politician. Observations are split into 50 bins with the `binscatter` command. Correlation values reported in the top-left corner for each panel.

Figure A6: Emotions



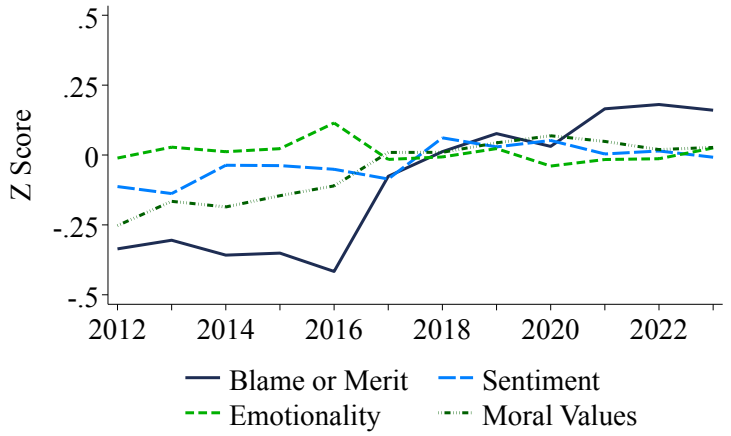
Notes: Each bar represents the share of tweets, within that rhetoric, classified as the corresponding emotion.

Figure A7: Supply of Blame and Merit Tweets over Time, Excluding Senators



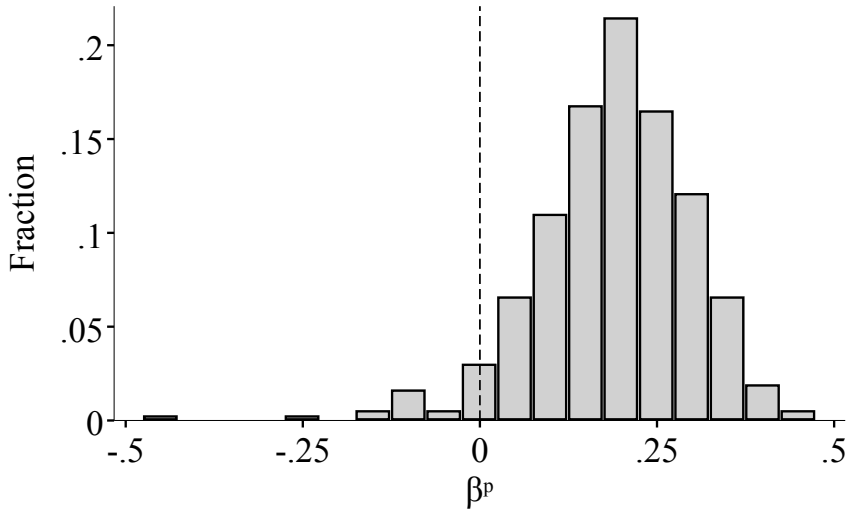
Notes: The figure presents the yearly share of tweets classified as blame and merit, excluding tweets from Senators. Shaded areas represent 95 percent confidence intervals.

Figure A8: Comparison with Other Text Measures



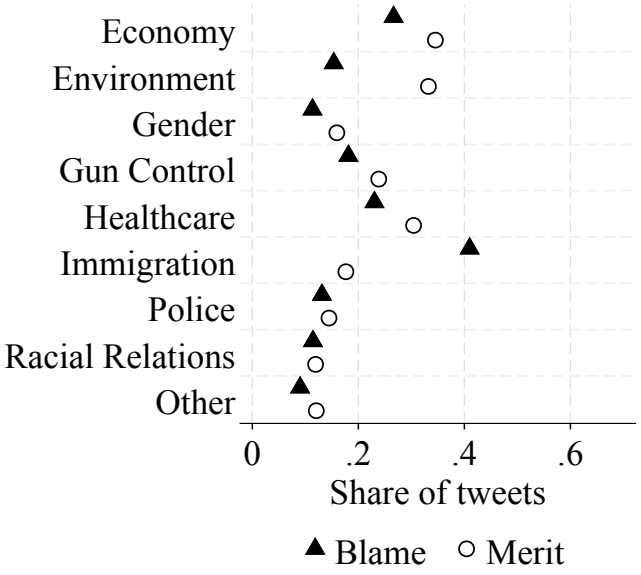
Notes: All text measures are standardized. Shaded areas represent 95 percent confidence intervals.

Figure A9: Individual Shifts to Causal Rhetoric



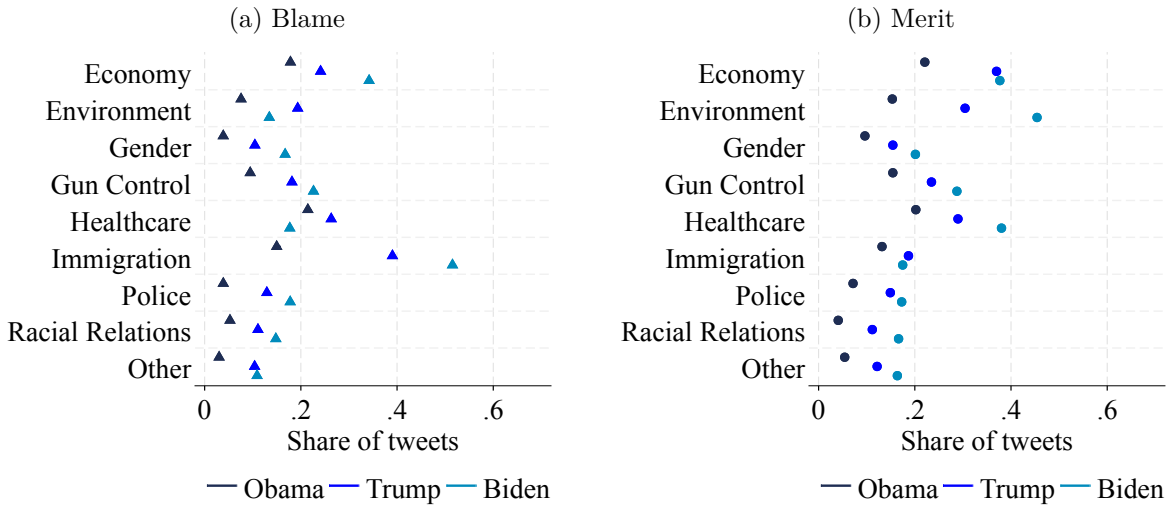
Notes: The coefficient reports the distribution of $\{\beta_i\}_i$ from the regression $y_{it} = \alpha + \beta_i D_{it}^{2017} + \varepsilon_{it}$ where y_{it} is binary indicator denoting whether tweet i is either blame or merit, D_{it}^{2017} is binary indicator taking the value 1 if tweet i is posted in or after 2017, estimated separately for each politician who appears in the dataset before and in or after 2017.

Figure A10: Blame and Merit over Topics



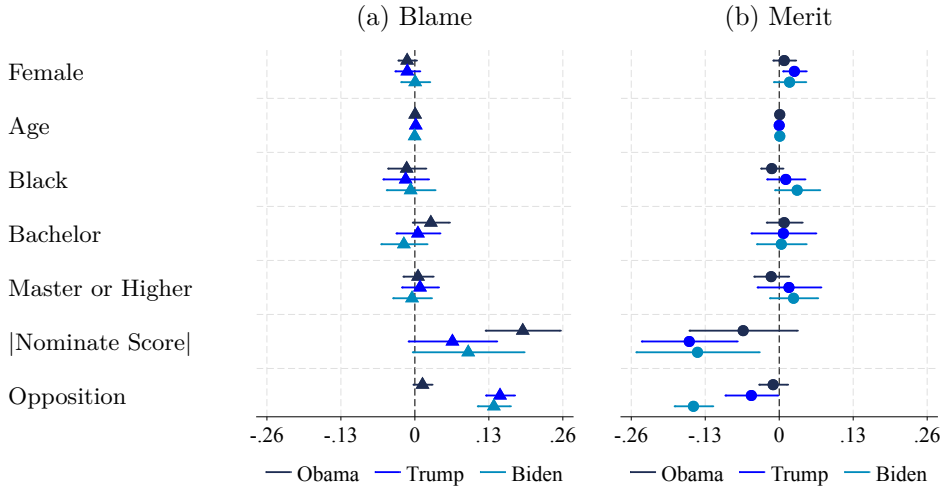
Notes: The figure presents the share of blame and merit tweets within each topic. Bars represent 95% confidence intervals.

Figure A11: Blame and Merit over Topics and Time



Notes: Both panels report the share of blame and merit tweets within each topic for each presidency. Bars represent 95% confidence intervals.

Figure A12: Author Correlates of Blame and Merit



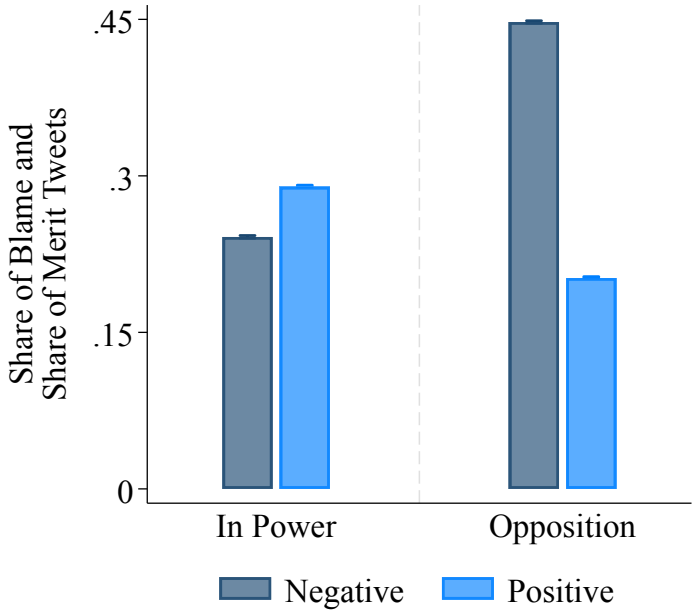
Notes: The coefficients presents the estimates from three regressions at the politician level of blame (merit) over the listed variables and the text-level measures of Figure 1, computed separately for each presidency. Bars represent 95 percent confidence interval with robust standard errors.

Figure A13: Rhetorical Adjustments over Time



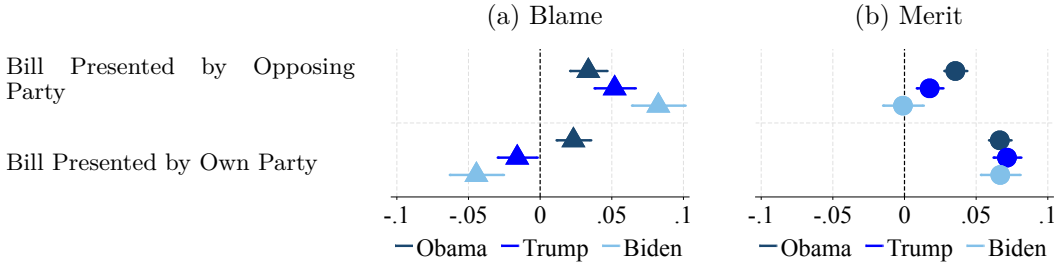
Notes: Both panels report coefficients from $y_{ipq} = \sum_{k=2012q2}^{2023q3} \beta^k D^k_{i,p,q} + \mu_p + \varepsilon_{ipq}$, where the outcome is either whether the tweet is blame (Panel (a)) or merit (Panel (b)) and μ_p denote politician fixed effects. Shaded areas represent 95 percent confidence intervals with standard errors clustered at the politician level.

Figure A14: Trade-Off and Sentiment



Notes: The figure presents the share of blame tweets among those expressing negative sentiment, and the share of merit tweets among those expressing positive sentiment, separately for politicians in power and in opposition. Bars represent 95% confidence intervals.

Figure A15: Legislative Activity: Robustness within Sentiment



Notes: The figure present the estimates of Equation 4, considering as outcomes a binary indicator for whether the tweet is blame and a binary indicator for whether the tweet is merit in Panels (a) and (b), respectively. In Panel (a), we restrict to tweets classified as having a negative sentiment, while in Panel (b) we restrict tweets classified as having a positive sentiment. Bars represent 95 percent confidence intervals computed with standard errors clustered at the day level.

Table A1: Descriptive Statistics

	Panel A) Tweet Level			*Panel B) Politician Level*		
	All	Democrats	Republicans	All	Democrats	Republicans
Female	0.300	0.401	0.165	0.243	0.378	0.132
	(0.458)	(0.490)	(0.371)	(0.429)	(0.486)	(0.339)
Age	57.973	59.058	56.509	57.468	58.499	56.617
	(11.572)	(12.075)	(10.682)	(11.767)	(12.649)	(10.926)
Black	0.082	0.127	0.021	0.080	0.150	0.022
	(0.274)	(0.333)	(0.142)	(0.271)	(0.357)	(0.148)
Bachelor	0.333	0.297	0.381	0.332	0.295	0.363
	(0.471)	(0.457)	(0.486)	(0.471)	(0.457)	(0.481)
Master or Higher	0.620	0.672	0.551	0.584	0.651	0.529
	(0.485)	(0.469)	(0.497)	(0.493)	(0.477)	(0.500)
Republican	0.426	0	1	0.548	0	1
	(0.494)	(0)	(0)	(0.498)	(0)	(0)
Nominate Score	-0.005	-0.387	0.511	0.109	-0.366	0.500
	(0.467)	(0.126)	(0.164)	(0.455)	(0.122)	(0.162)
Representative	0.804	0.814	0.790	0.857	0.857	0.856
	(0.397)	(0.389)	(0.407)	(0.351)	(0.350)	(0.351)
Senator				0.118	0.123	0.114
				(0.323)	(0.329)	(0.318)
Both Representative and Senator				0.026	0.020	0.030
				(0.158)	(0.139)	(0.172)
Number of Tweets Posted				4664.950	5924.391	3625.209
				(4848.342)	(5151.747)	(4319.150)
Number of Accounts				1.988	2.042	1.943
				(0.689)	(0.688)	(0.687)
Observations	4198455	2411227	1787228	900	407	493

Notes: Panel A presents statitics at the tweet level. Panel B presents statistics at the politician level. Means and standard deviations in parentheses.

Table A2: Labeled Tweets

Tweet	Causal	Tone	Rhetoric Style
Biden has lost all credibility. No one believes your lies, Joe! URL	0	-1	None
Juan Williams just compared Obamacare to a sweater. Good analogy. It's something you don't want to get, but have to accept when given.	0	-1	None
Joe Biden and the Democrats terrible policies have wreaked havoc on this country.	1	-1	Blame
#Trumpcare is fundamentally flawed. Higher costs, less coverage, fewer protections – thats GOPs gift to American people. #ProtectOurCare URL	1	-1	Blame
I couldnt support final passage of todays approps package but Im pleased about the inclusion of my HBCU amendment URL	0	1	None
Universal congrats to the scientists at @OregonState for their work helping Insight make a successful #MarsLanding: URL	0	1	None
The Protect Medical Innovation Act will boost American innovation and manufacturing, and it will encourage medical research and development that make a real difference in peoples lives. URL	1	1	Merit
Now that the Inflation Reduction Act is law it will not only lower prescription drug prices but save lives. Thank you @HenryFordHealth for your support. URL	1	1	Merit

Table A3: Democrats Tweets Classified as Merit and Blame

Panel (a) Merit Tweets
Proud to cosponsor @NydiaVelazquez's Public Housing Emergency Response Act which would invest $70B in public housing including $32B for @NYCHA. Our public housing crisis must be addressed & this bill is a bold approach to doing that. Residents deserve to live in safe conditions!
I applaud @POTUS for setting our next offshore wind target. With a new infusion of investments from my offshore wind manufacturing tax credit in the Inflation Reduction Act, the U.S. can and will deploy 15 GW of floating offshore wind by 2035 all while creating good union jobs. URL QT @ginamccarthy46 Today we're announcing actions to advance *floating* offshore wind platforms – key to harnessing the potential of deep waters along the West Coast, Gulf of Maine, & more. Part of @POTUS' plan for American jobs and leadership on new clean energy technologies! URL
The Bipartisan Infrastructure Law is putting our economy on track to thrive & investing in communities that have too often been left behind. With over $2 million recently announced, we take a major step toward redeveloping Baltimores Bridge to Nowhere. URL
Happy to be joining @HouseDemocrats to help Americas workers access better paying jobs. The Workforce Innovation and Opportunity Act connects employers with qualified candidates, lowers costs for families and increases supplies. Democrats are #InvestingInWorkers. URL
The #BuildBackBetterAct provides much needed funds to @TheJusticeDept to help reduce community violence & fund proven intervention programs. Im proud to advocate for legislation to break cycles of violence in communities, saving American lives & taxpayer dollars.
Panel (b) Blame Tweets
We warned when the GOP passed tax cuts for the rich that it would explode deficits. It did. We warned that the GOP would use those deficits to come after Social Security and Medicare. They are. URL
The party of NO, #ILGOP in particular, plays political games in ignoring the implications on our economy, on jobs, Social Security checks. Republicans raised the debt ceiling 3 times under Trump's thumb. They are playing politics with people's lives. URL
GOPs reckless health care strategy is already destabilizing #healthcare markets and forcing premiums to rise.
If #GOP could govern, we wouldn't be having this conversation. The root of #TrumpShutdown is incompetence, putting veterans and kids health care at risk, pensions at risk, and the fight against opioids at risk
Real wages today are lower than they were in 1973. Thats not a sign of a healthy economy, its a sign that working people today are worse off than they were 45 years ago, and the GOP tax cuts have done nothing to address that issue.

Table A4: Republicans Tweets Classified as Merit and Blame

Panel (a) Merit Tweets
I'm an original sponsor of the Nuclear Energy Leadership Act w/ @RepElaineLuria to encourage further dev of advanced nuclear energy programs. Such programs will create high-quality jobs, strengthen natl security, reduce foreign energy dependence and promote emissions-free energy.
Today @realDonaldTrump showed his commitment to supporting American energy dominance. The @EPA's rule will bolster our nations energy independence by lowering energy costs, spurring job growth and promoting economic development in our communities. URL
Glad to hear @realDonaldTrump has signed legislation adding $320 billion to the #PaycheckProtectionProgram, ramping up testing capability and providing more funding for health care providers. AR will benefit from this measure to protect public health and save businesses & jobs. URL URL
Earlier this week, I introduced #CARA2 to increase funding levels for programs we know work and implements additional policy reforms that will make a real difference in combatting the #opioidcrisis. URL
@TransportGOP are delivering on our promise of fixing supply chain holes and building a stronger economy. Currently we are marking up a package of bills that will remove barriers, increase efficiency, and target infrastructure investment. #SupplyChain
Panel (b) Blame Tweets
The supply chain and inflation crises are not a high class problem like @WHCOS claims. As Dems look to pour trillions into the economy and spike inflation further, they must understand that actions have consequences that will be felt by every American URL
Our country is facing soaring inflation thanks to Democrats spending spree, and whats @POTUS solution? Spend MORE money.
@JoeBiden and @SenateDems are TOTALLY out of touch with reality. Inflation is still wiping out wage growth, all while Democrats' reckless spending spree makes matters worse. URL
Top border officials told Biden that if he unraveled Trump's policies and pushed for open borders that a major crisis would occur. He didnt listen. Now everyone is suffering – Americans and migrants alike. URL
April saw the highest number of migrants ever recorded. Next week, @JoeBiden will reverse another commonsense border policy that will only make this crisis worse. Biden needs to wake up and face reality. URL

Table A5: List of Control Variables for Additional Interactions

Variable
Population density
Log(County area)
Distance from Austin, TX (in miles)
Distance from NYC (in miles)
Distance from San Francisco (in miles)
Distance from Washington, DC (in miles)
% aged 20–24
% aged 25–29
% aged 30–34
% aged 35–39
% aged 40–44
% aged 45–49
% aged 50+
Population growth, 2000–2016
% white
% black
% native American
% Asian
% Hispanic
% unemployed
% below poverty level
% employed in IT
% employed in construction/real estate
% employed in manufacturing
% with high school degree
% with college education
% watching Fox News
% watching prime time TV

Notes: The table presents the cross-sectional county-level controls that we interact with the shares of blame and merit tweets in Equations 3 and 5.

Table A6: Donations

	(1)	(2)	(3)	(4)
	Panel A) Revenue from Donations			
Blame x Users	0.0074***	0.0084***	0.0055***	0.0064**
	(0.0013)	(0.0029)	(0.0011)	(0.0028)
Merit x Users	0.0007*	0.0015	0.0010**	0.0018*
	(0.0004)	(0.0010)	(0.0004)	(0.0010)
Blame x SXSWFollower2006	-0.0077	-0.0050	-0.0072*	-0.0044
	(0.0050)	(0.0043)	(0.0042)	(0.0036)
Merit x SXSWFollower2006	0.0030*	0.0023	0.0030	0.0022
	(0.0018)	(0.0016)	(0.0019)	(0.0016)
	Panel B) Number of Donors			
Blame x Users	0.0038***	0.0051***	0.0029***	0.0042***
	(0.0006)	(0.0014)	(0.0005)	(0.0013)
Merit x Users	-0.0000	-0.0000	0.0001	0.0001
	(0.0002)	(0.0004)	(0.0002)	(0.0004)
Blame x SXSWFollower2006	-0.0026	-0.0013	-0.0027	-0.0013
	(0.0023)	(0.0021)	(0.0019)	(0.0018)
Merit x SXSWFollower2006	0.0010	0.0006	0.0010	0.0006
	(0.0007)	(0.0005)	(0.0007)	(0.0005)
	Panel C) Average per Donor			
Blame x Users	0.0042***	0.0037**	0.0029***	0.0025
	(0.0008)	(0.0019)	(0.0007)	(0.0018)
Merit x Users	0.0006*	0.0015*	0.0008**	0.0016**
	(0.0003)	(0.0008)	(0.0003)	(0.0008)
Blame x SXSWFollower2006	-0.0054	-0.0040	-0.0047*	-0.0033
	(0.0033)	(0.0028)	(0.0028)	(0.0023)
Merit x SXSWFollower2006	0.0023*	0.0020	0.0022*	0.0019
	(0.0013)	(0.0012)	(0.0013)	(0.0012)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107
F statistic	95.47	31.80	63.65	21.20
Partial F statistic Blame x User	190.95	63.60	190.95	63.60
Partial F statistic Merit x User	190.95	63.60	190.95	63.60

Notes: The table presents the 2SLS estimates of Equation 3. In Panel A) the outcome is the log+1 of the revenue from donations. In Panel B) the outcome is the log+1 of the number of donors. In Panel C) the outcome is the log+1 of the average amount donated per donor. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of Twitter users in the county. We also control for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. The interactions with Twitter users are instrumented using the corresponding interaction with SXSW followers in the county who joined in 2007. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A7: Donations OLS

	(1)	(2)	(3)	(4)
	Panel A) Revenue from Donations			
Blame x Users	0.0076***	0.0086***	0.0062***	0.0072***
	(0.0003)	(0.0004)	(0.0002)	(0.0004)
Merit x Users	-0.0007***	-0.0013***	-0.0005***	-0.0012***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
	Panel B) Number of Donors			
Blame x Users	0.0031***	0.0034***	0.0026***	0.0029***
	(0.0001)	(0.0002)	(0.0001)	(0.0002)
Merit x Users	-0.0003***	-0.0005***	-0.0003***	-0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	Panel C) Average per Donor			
Blame x Users	0.0052***	0.0060***	0.0042***	0.0050***
	(0.0002)	(0.0003)	(0.0002)	(0.0003)
Merit x Users	-0.0005***	-0.0010***	-0.0004***	-0.0009***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107

Notes: The table presents the OLS estimates of Equation 3. In Panel A) the outcome is the log+1 of the revenue from donations. In Panel B) the outcome is the log+1 of the number of donors. In Panel C) the outcome is the log+1 of the average amount donated per donor. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of Twitter users in the county. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A8: Donations Reduced Form

	(1)	(2)	(3)	(4)
	Panel A) Revenue from Donations			
Blame x SXSWFollower2007	0.0185***	0.0086***	0.0136***	0.0038
	(0.0035)	(0.0032)	(0.0030)	(0.0028)
Merit x SXSWFollower2007	0.0019*	0.0016*	0.0024**	0.0022**
	(0.0011)	(0.0010)	(0.0011)	(0.0010)
Blame x SXSWFollower2006	-0.0097	-0.0097*	-0.0087	-0.0086**
	(0.0066)	(0.0054)	(0.0054)	(0.0044)
Merit x SXSWFollower2006	0.0028	0.0014	0.0027	0.0013
	(0.0019)	(0.0016)	(0.0019)	(0.0017)
	Panel B) Number of Donors			
Blame x SXSWFollower2007	0.0094***	0.0052***	0.0073***	0.0031**
	(0.0016)	(0.0015)	(0.0014)	(0.0013)
Merit x SXSWFollower2007	-0.0000	-0.0000	0.0002	0.0002
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Blame x SXSWFollower2006	-0.0037	-0.0042	-0.0035	-0.0040*
	(0.0031)	(0.0026)	(0.0025)	(0.0021)
Merit x SXSWFollower2006	0.0010	0.0006	0.0009	0.0006
	(0.0007)	(0.0006)	(0.0007)	(0.0006)
	Panel C) Average per Donor			
Blame x SXSWFollower2007	0.0104***	0.0038*	0.0073***	0.0007
	(0.0022)	(0.0020)	(0.0019)	(0.0018)
Merit x SXSWFollower2007	0.0016*	0.0015**	0.0019**	0.0019**
	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Blame x SXSWFollower2006	-0.0065	-0.0061*	-0.0055	-0.0051*
	(0.0042)	(0.0034)	(0.0034)	(0.0028)
Merit x SXSWFollower2006	0.0021	0.0011	0.0020	0.0010
	(0.0014)	(0.0013)	(0.0014)	(0.0013)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107

Notes: The table presents the reduced form estimates of Equation 3. In Panel A) the outcome is the log+1 of the revenue from donations. In Panel B) the outcome is the log+1 of the number of donors. In Panel C) the outcome is the log+1 of the average amount donated per donor. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of SXSW followers in the county who joined in 2007. We also control for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A9: Donations First Stage

	(1)	(2)	(3)	(4)
	Panel A) Blame x Users			
Blame x SXSWFollower2007	2.4824***	1.0283***	2.4824***	1.0283***
	(0.1796)	(0.1289)	(0.1796)	(0.1289)
Merit x SXSWFollower2007	-0.0000***	0.0000	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Blame x SXSWFollower2006	-0.2744	-0.5627**	-0.2744	-0.5627**
	(0.3938)	(0.2450)	(0.3938)	(0.2450)
Merit x SXSWFollower2006	0.0000	-0.0000	0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	Panel B) Merit x Users			
Blame x SXSWFollower2007	-0.0000***	0.0000	-0.0000***	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Merit x SXSWFollower2007	2.4824***	1.0283***	2.4824***	1.0283***
	(0.1796)	(0.1289)	(0.1796)	(0.1289)
Blame x SXSWFollower2006	0.0000	-0.0000	0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Merit x SXSWFollower2006	-0.2744	-0.5627**	-0.2744	-0.5627**
	(0.3938)	(0.2450)	(0.3938)	(0.2450)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107

Notes: The table presents the first-stage regressions related to the 2SLS estimation of Equation 3. In Panel A) the outcome is the monthly share of blame tweets posted by the politician interacted with the log+1 number of Twitter users in the county. In Panel B) the outcome is the monthly share of merit tweets posted by the politician interacted with the log+1 number of Twitter users in the county. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of SXSW followers in the county who joined in 2007. The table also includes the estimates for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A10: Donations from Moderate Donors

	(1)	(2)	(3)	(4)
	Panel A) Revenue from Donations			
Blame x Users	-0.0018***	-0.0011	-0.0030***	-0.0023*
	(0.0004)	(0.0011)	(0.0006)	(0.0012)
Merit x Users	0.0044***	0.0055***	0.0045***	0.0056***
	(0.0006)	(0.0015)	(0.0006)	(0.0015)
Blame x SXSWFollower2006	-0.0009	0.0001	-0.0011	-0.0001
	(0.0020)	(0.0017)	(0.0028)	(0.0024)
Merit x SXSWFollower2006	-0.0000	0.0003	0.0000	0.0003
	(0.0029)	(0.0027)	(0.0030)	(0.0028)
	Panel B) Number of Donors			
Blame x Users	-0.0011***	-0.0013***	-0.0017***	-0.0018***
	(0.0002)	(0.0005)	(0.0002)	(0.0005)
Merit x Users	0.0020***	0.0027***	0.0021***	0.0028***
	(0.0003)	(0.0006)	(0.0003)	(0.0007)
Blame x SXSWFollower2006	-0.0001	-0.0001	-0.0003	-0.0003
	(0.0008)	(0.0008)	(0.0012)	(0.0011)
Merit x SXSWFollower2006	-0.0002	0.0002	-0.0001	0.0002
	(0.0012)	(0.0012)	(0.0013)	(0.0012)
	Panel C) Average per Donor			
Blame x Users	-0.0005	0.0004	-0.0013***	-0.0004
	(0.0003)	(0.0008)	(0.0004)	(0.0008)
Merit x Users	0.0024***	0.0028***	0.0025***	0.0029***
	(0.0004)	(0.0010)	(0.0004)	(0.0010)
Blame x SXSWFollower2006	-0.0005	0.0004	-0.0006	0.0003
	(0.0012)	(0.0010)	(0.0017)	(0.0014)
Merit x SXSWFollower2006	0.0002	0.0003	0.0002	0.0003
	(0.0018)	(0.0017)	(0.0019)	(0.0017)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107
F statistic	95.47	31.80	63.65	21.20
Partial F statistic Blame x User	190.95	63.60	190.95	63.60
Partial F statistic Merit x User	190.95	63.60	190.95	63.60

Notes: The table presents the 2SLS estimates of Equation 3 restricting to donations from moderate donors. In Panel A) the outcome is the log+1 of the revenue from donations. In Panel B) the outcome is the log+1 of the number of donors. In Panel C) the outcome is the log+1 of the average amount donated per donor. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of Twitter users in the county. We also control for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. The interactions with Twitter users are instrumented using the corresponding interaction with SXSW followers in the county who joined in 2007. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A11: Donations from Extreme Donors

	(1)	(2)	(3)	(4)
	Panel A) Revenue from Donations			
Blame x Users	0.0123***	0.0145***	0.0110***	0.0133***
	(0.0018)	(0.0042)	(0.0017)	(0.0041)
Merit x Users	-0.0040***	-0.0048***	-0.0039***	-0.0046***
	(0.0006)	(0.0015)	(0.0006)	(0.0015)
Blame x SXSWFollower2006	-0.0063	-0.0041	-0.0054	-0.0032
	(0.0075)	(0.0069)	(0.0070)	(0.0064)
Merit x SXSWFollower2006	0.0031	0.0019	0.0030	0.0018
	(0.0025)	(0.0023)	(0.0024)	(0.0022)
	Panel B) Number of Donors			
Blame x Users	0.0056***	0.0074***	0.0050***	0.0069***
	(0.0008)	(0.0018)	(0.0007)	(0.0018)
Merit x Users	-0.0020***	-0.0027***	-0.0019***	-0.0026***
	(0.0003)	(0.0007)	(0.0003)	(0.0007)
Blame x SXSWFollower2006	-0.0024	-0.0011	-0.0022	-0.0010
	(0.0032)	(0.0030)	(0.0030)	(0.0028)
Merit x SXSWFollower2006	0.0010	0.0004	0.0010	0.0004
	(0.0011)	(0.0011)	(0.0011)	(0.0010)
	Panel C) Average per Donor			
Blame x Users	0.0075***	0.0078***	0.0067***	0.0070**
	(0.0012)	(0.0028)	(0.0011)	(0.0027)
Merit x Users	-0.0024***	-0.0025**	-0.0023***	-0.0024**
	(0.0004)	(0.0010)	(0.0004)	(0.0010)
Blame x SXSWFollower2006	-0.0043	-0.0034	-0.0035	-0.0026
	(0.0050)	(0.0045)	(0.0047)	(0.0042)
Merit x SXSWFollower2006	0.0023	0.0016	0.0022	0.0015
	(0.0016)	(0.0015)	(0.0016)	(0.0015)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107
F statistic	95.47	31.80	63.65	21.20
Partial F statistic Blame x User	190.95	63.60	190.95	63.60
Partial F statistic Merit x User	190.95	63.60	190.95	63.60

Notes: The table presents the 2SLS estimates of Equation 3 restricting to donations from extreme donors. In Panel A) the outcome is the log+1 of the revenue from donations. In Panel B) the outcome is the log+1 of the number of donors. In Panel C) the outcome is the log+1 of the average amount donated per donor. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of Twitter users in the county. We also control for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. The interactions with Twitter users are instrumented using the corresponding interaction with SXSW followers in the county who joined in 2007. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A12: Revenue from Donations over Time

	(1)	(2)	(3)	(4)
(Blame x Users) x Obama	0.0010**	0.0019	0.0003	0.0012
	(0.0004)	(0.0018)	(0.0004)	(0.0018)
(Blame x Users) x Trump	0.0023**	0.0032	0.0026**	0.0035
	(0.0010)	(0.0026)	(0.0012)	(0.0027)
(Blame x Users) x Biden	0.0149***	0.0158***	0.0088***	0.0098***
	(0.0022)	(0.0038)	(0.0015)	(0.0031)
(Merit x Users) x Obama	0.0033***	0.0041***	0.0033***	0.0041***
	(0.0006)	(0.0010)	(0.0006)	(0.0010)
(Merit x Users) x Trump	0.0046***	0.0054***	0.0042***	0.0050***
	(0.0007)	(0.0011)	(0.0006)	(0.0010)
(Merit x Users) x Biden	-0.0045***	-0.0037**	-0.0026***	-0.0018
	(0.0012)	(0.0015)	(0.0010)	(0.0014)
(Blame x SXSWFollower2006) x Obama	-0.0005	0.0022	0.0006	0.0033
	(0.0017)	(0.0030)	(0.0019)	(0.0031)
(Blame x SXSWFollower2006) x Trump	-0.0038	-0.0011	-0.0060	-0.0033
	(0.0036)	(0.0033)	(0.0043)	(0.0039)
(Blame x SXSWFollower2006) x Biden	-0.0148*	-0.0120	-0.0121**	-0.0094*
	(0.0089)	(0.0081)	(0.0057)	(0.0049)
(Merit x SXSWFollower2006) x Obama	0.0014	0.0007	0.0013	0.0006
	(0.0026)	(0.0026)	(0.0026)	(0.0026)
(Merit x SXSWFollower2006) x Trump	-0.0005	-0.0012	0.0001	-0.0006
	(0.0030)	(0.0031)	(0.0028)	(0.0029)
(Merit x SXSWFollower2006) x Biden	0.0068	0.0061	0.0059	0.0051
	(0.0046)	(0.0042)	(0.0039)	(0.0034)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107
F statistic	31.82	10.60	21.22	7.07

Notes: The table presents the 2SLS estimates of Equation 3 allowing the coefficients of the terms involving the share of blame and the share of merit tweets to vary across presidencies. The outcome is the log+1 of the revenue from donations. The outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of Twitter users in the county. We also include controls for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. To allow for heterogeneity across presidencies, all interactions are further interacted with presidency indicators. The interactions with Twitter users are instrumented using the corresponding interaction with SXSW followers in the county who joined in 2007. Standard errors in parentheses clustered at county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A13: Number of Donors over Time

	(1)	(2)	(3)	(4)
(Blame x Users) x Obama	-0.0001	0.0012	-0.0004***	0.0009
	(0.0001)	(0.0008)	(0.0001)	(0.0008)
(Blame x Users) x Trump	0.0012***	0.0024**	0.0014**	0.0027**
	(0.0004)	(0.0012)	(0.0005)	(0.0013)
(Blame x Users) x Biden	0.0080***	0.0093***	0.0054***	0.0067***
	(0.0010)	(0.0018)	(0.0007)	(0.0015)
(Merit x Users) x Obama	0.0009***	0.0009***	0.0009***	0.0009***
	(0.0001)	(0.0003)	(0.0001)	(0.0003)
(Merit x Users) x Trump	0.0022***	0.0021***	0.0020***	0.0019***
	(0.0003)	(0.0004)	(0.0003)	(0.0004)
(Merit x Users) x Biden	-0.0025***	-0.0025***	-0.0016***	-0.0017***
	(0.0005)	(0.0007)	(0.0004)	(0.0006)
(Blame x SXSWFollower2006) x Obama	0.0004	0.0017	0.0005	0.0019
	(0.0005)	(0.0014)	(0.0006)	(0.0014)
(Blame x SXSWFollower2006) x Trump	-0.0014	-0.0001	-0.0025	-0.0012
	(0.0016)	(0.0015)	(0.0020)	(0.0019)
(Blame x SXSWFollower2006) x Biden	-0.0052	-0.0039	-0.0043	-0.0030
	(0.0043)	(0.0040)	(0.0029)	(0.0027)
(Merit x SXSWFollower2006) x Obama	0.0005	0.0001	0.0005	0.0001
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
(Merit x SXSWFollower2006) x Trump	-0.0005	-0.0009	-0.0003	-0.0007
	(0.0013)	(0.0013)	(0.0012)	(0.0012)
(Merit x SXSWFollower2006) x Biden	0.0024	0.0020	0.0020	0.0016
	(0.0019)	(0.0017)	(0.0016)	(0.0013)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107
F statistic	31.82	10.60	21.22	7.07

Notes: The table presents the 2SLS estimates of Equation 3 allowing the coefficients of the terms involving the share of blame and the share of merit tweets to vary across presidencies. The outcome is the log+1 of the number of donors. The outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of Twitter users in the county. We also include controls for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. To allow for heterogeneity across presidencies, all interactions are further interacted with presidency indicators. The interactions with Twitter users are instrumented using the corresponding interaction with SXSW followers in the county who joined in 2007. Standard errors in parentheses clustered at county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A14: Average per Donor over Time

	(1)	(2)	(3)	(4)
(Blame x Users) x Obama	0.0014***	0.0009	0.0009**	0.0004
	(0.0003)	(0.0013)	(0.0004)	(0.0013)
(Blame x Users) x Trump	0.0015**	0.0010	0.0012*	0.0008
	(0.0007)	(0.0017)	(0.0007)	(0.0018)
(Blame x Users) x Biden	0.0075***	0.0070***	0.0041***	0.0037*
	(0.0014)	(0.0024)	(0.0010)	(0.0020)
(Merit x Users) x Obama	0.0026***	0.0035***	0.0026***	0.0035***
	(0.0005)	(0.0008)	(0.0005)	(0.0008)
(Merit x Users) x Trump	0.0025***	0.0033***	0.0023***	0.0032***
	(0.0005)	(0.0008)	(0.0004)	(0.0008)
(Merit x Users) x Biden	-0.0027***	-0.0018	-0.0017**	-0.0008
	(0.0008)	(0.0011)	(0.0007)	(0.0010)
(Blame x SXSWFollower2006) x Obama	-0.0014	-0.0000	-0.0003	0.0011
	(0.0015)	(0.0021)	(0.0016)	(0.0022)
(Blame x SXSWFollower2006) x Trump	-0.0023	-0.0009	-0.0036	-0.0023
	(0.0025)	(0.0023)	(0.0028)	(0.0025)
(Blame x SXSWFollower2006) x Biden	-0.0100*	-0.0086*	-0.0078**	-0.0064**
	(0.0055)	(0.0048)	(0.0037)	(0.0031)
(Merit x SXSWFollower2006) x Obama	0.0008	0.0005	0.0007	0.0004
	(0.0021)	(0.0021)	(0.0021)	(0.0021)
(Merit x SXSWFollower2006) x Trump	-0.0001	-0.0004	0.0003	0.0000
	(0.0019)	(0.0020)	(0.0018)	(0.0019)
(Merit x SXSWFollower2006) x Biden	0.0054	0.0051	0.0046	0.0043
	(0.0034)	(0.0032)	(0.0029)	(0.0027)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	168,232,932	168,178,803	168,232,932	168,178,803
Clusters	3,108	3,107	3,108	3,107
F statistic	31.82	10.60	21.22	7.07

Notes: The table presents the 2SLS estimates of Equation 3 allowing the coefficients of the terms involving the share of blame and the share of merit tweets to vary across presidencies. The outcome is the log+1 of the average amount donated per donor. The outcome is regressed on the monthly share of blame and merit tweets posted by the politician interacted with the log+1 number of Twitter users in the county. We also include controls for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. To allow for heterogeneity across presidencies, all interactions are further interacted with presidency indicators. The interactions with Twitter users are instrumented using the corresponding interaction with SXSW followers in the county who joined in 2007. Standard errors in parentheses clustered at county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A15: Legislative Activity

	Tweet is Blame	Tweet is Merit
Bill Presented by Opposing Party x Obama	0.013***	0.023***
	(0.003)	(0.003)
Bill Presented by Opposing Party x Trump	0.051***	0.007
	(0.005)	(0.004)
Bill Presented by Opposing Party x Biden	0.069***	-0.014**
	(0.007)	(0.007)
Bill Presented by Own Party x Obama	0.026***	0.047***
	(0.003)	(0.003)
Bill Presented by Own Party x Trump	-0.011**	0.061***
	(0.005)	(0.004)
Bill Presented by Own Party x Biden	-0.025***	0.059***
	(0.007)	(0.007)
Politician FE	✓	✓
Week FE	✓	✓
Observations	4,198,452	4,198,452
Clusters	4,202	4,202

Notes: The table presents the estimates of Equation 4. In the first column the outcome is a binary indicator equal to 1 if the tweet is classified as blame and 0 otherwise. In the second column the outcome is a binary indicator equal to 1 if the tweet is classified as merit and 0 otherwise. In each column, the outcome is regressed on two binary indicators equal to 1 if a bill is presented by a member of the opposing party on that day and 0 otherwise, and equal to 1 if a bill is presented by a member of the own party on that day and 0 otherwise. To allow for heterogeneity across presidencies, these indicators are interacted with presidency indicators. Standard errors in parentheses clustered at the day level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A16: Legislative Activity: Robustness within Sentiment

	Tweet is Blame	Tweet is Merit
Bill Presented by Opposing Party	0.034***	0.035***
	(0.006)	(0.004)
Bill Presented by Opposing Party x Trump	0.052***	0.018***
	(0.007)	(0.005)
Bill Presented by Opposing Party x Biden	0.082***	-0.001
	(0.009)	(0.007)
Bill Presented by Own Party	0.023***	0.066***
	(0.006)	(0.004)
Bill Presented by Own Party x Trump	-0.016**	0.072***
	(0.007)	(0.005)
Bill Presented by Own Party x Biden	-0.044***	0.067***
	(0.009)	(0.007)
Politician FE	✓	✓
Week FE	✓	✓
Observations	985,824	2,610,253
Clusters	4,202	4,202

Notes: The table presents the estimates of Equation 4. In the first column the outcome is a binary indicator equal to 1 if the tweet is classified as blame and 0 otherwise, with the estimation restricted to tweets classified as having a negative sentiment. In the second column the outcome is a binary indicator equal to 1 if the tweet is classified as merit and 0 otherwise, with the estimation restricted to tweets classified as having a positive sentiment. In each column, the outcome is regressed on two binary indicators equal to 1 if a bill is presented by a member of the opposing party on that day and 0 otherwise, and equal to 1 if a bill is presented by a member of the same party on that day and 0 otherwise. To allow for heterogeneity across presidencies, these indicators are interacted with presidency indicators. Standard errors in parentheses clustered at the day level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A17: Protests

	(1)	(2)	(3)	(4)
	Panel A) Any Protest			
Blame x Users	0.0083***	0.0184***	0.0090***	0.0190***
	(0.0023)	(0.0064)	(0.0030)	(0.0067)
Merit x Users	0.0028	-0.0117	0.0023	-0.0122*
	(0.0030)	(0.0073)	(0.0031)	(0.0072)
Blame x SXSWFollower2006	-0.0194**	-0.0138*	-0.0207*	-0.0151
	(0.0093)	(0.0072)	(0.0121)	(0.0098)
Merit x SXSWFollower2006	0.0027	-0.0168	0.0037	-0.0158
	(0.0117)	(0.0105)	(0.0124)	(0.0109)
	Panel B) Number of Protests			
Blame x Users	0.0138***	0.0343***	0.0158***	0.0363***
	(0.0042)	(0.0106)	(0.0050)	(0.0110)
Merit x Users	0.0432***	0.0452***	0.0417***	0.0437***
	(0.0075)	(0.0139)	(0.0077)	(0.0139)
Blame x SXSWFollower2006	0.0164	0.0315**	0.0143	0.0293
	(0.0176)	(0.0147)	(0.0212)	(0.0184)
Merit x SXSWFollower2006	0.0635**	0.0276	0.0651**	0.0292
	(0.0277)	(0.0268)	(0.0275)	(0.0267)
Politician FE	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	245,532	245,453	245,532	245,453
Clusters	3,108	3,107	3,108	3,107
F statistic	95.47	31.79	63.65	21.19
Partial F statistic Blame x User	190.99	63.52	190.95	63.52
Partial F statistic Merit x User	190.99	63.60	190.95	63.60

Notes: The table presents the 2SLS estimates of Equation 5. In Panel A) the outcome is a binary indicator equal to 1 if at least one protest and 0 otherwise. In Panel B) the outcome is the log+1 of the number of protest. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by all politicians interacted with the log+1 number of Twitter users in the county. We also control for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. The interactions with Twitter users are instrumented using the corresponding interaction with SXSW followers in the county who joined in 2007. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A18: Protests OLS

	(1)	(2)	(3)	(4)
	Panel A) Any Protest			
Blame x Users	0.0001	0.0014	0.0015**	0.0028***
	(0.0006)	(0.0010)	(0.0007)	(0.0010)
Merit x Users	0.0066***	0.0041***	0.0055***	0.0030***
	(0.0006)	(0.0010)	(0.0006)	(0.0010)
	Panel B) Number of Protests			
Blame x Users	0.0015	0.0018	0.0028***	0.0031**
	(0.0009)	(0.0014)	(0.0011)	(0.0015)
Merit x Users	0.0253***	0.0179***	0.0243***	0.0169***
	(0.0016)	(0.0018)	(0.0016)	(0.0018)
Politician FE	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	245,532	245,453	245,532	245,453
Clusters	3,108	3,107	3,108	3,107

Notes: The table presents the OLS estimates of Equation 5. In Panel A) the outcome is a binary indicator equal to 1 if at least one protest and 0 otherwise. In Panel B) the outcome is the log+1 of the number of protest. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by all politicians interacted with the log+1 number of Twitter users in the county. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A19: Protests Reduced Form

	(1)	(2)	(3)	(4)
	Panel A) Any Protest			
Blame x SXSWFollower2007	0.0207***	0.0189***	0.0223***	0.0204***
	(0.0056)	(0.0063)	(0.0073)	(0.0078)
Merit x SXSWFollower2007	0.0069	-0.0120	0.0058	-0.0132*
	(0.0075)	(0.0074)	(0.0078)	(0.0075)
Blame x SXSWFollower2006	-0.0217**	-0.0241**	-0.0232*	-0.0256**
	(0.0095)	(0.0094)	(0.0126)	(0.0123)
Merit x SXSWFollower2006	0.0019	-0.0103	0.0030	-0.0091
	(0.0128)	(0.0125)	(0.0134)	(0.0128)
	Panel B) Number of Protests			
Blame x SXSWFollower2007	0.0342***	0.0352***	0.0393***	0.0403***
	(0.0096)	(0.0100)	(0.0118)	(0.0123)
Merit x SXSWFollower2007	0.1073***	0.0465***	0.1035***	0.0427***
	(0.0205)	(0.0151)	(0.0204)	(0.0151)
Blame x SXSWFollower2006	0.0126	0.0122	0.0099	0.0095
	(0.0170)	(0.0165)	(0.0208)	(0.0206)
Merit x SXSWFollower2006	0.0517	0.0022	0.0537	0.0042
	(0.0342)	(0.0327)	(0.0340)	(0.0327)
Politician FE	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	245,532	245,453	245,532	245,453
Clusters	3,108	3,107	3,108	3,107
F statistic	0.51	0.51	0.51	0.51
Partial F statistic Blame x User	190.99	63.52	190.95	63.52
Partial F statistic Merit x User	190.99	63.60	190.95	63.60

Notes: The table presents the reduced form estimates of Equation 5. In Panel A) the outcome is a binary indicator equal to 1 if at least one protest and 0 otherwise. In Panel B) the outcome is the log+1 of the number of protest. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by all politicians interacted with the log+1 number of SXSW followers in the county who joined in 2007. We also control for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A20: Protests First Stage

	(1)	(2)	(3)	(4)
	Panel A) Blame x Users			
Blame x SXSWFollower2007	2.4824***	1.0283***	2.4824***	1.0283***
	(0.1796)	(0.1290)	(0.1796)	(0.1290)
Merit x SXSWFollower2007	-0.0000***	-0.0000	-0.0000***	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Blame x SXSWFollower2006	-0.2744	-0.5627**	-0.2744	-0.5627**
	(0.3938)	(0.2451)	(0.3938)	(0.2451)
Merit x SXSWFollower2006	0.0000	-0.0000	0.0000***	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	Panel B) Merit x Users			
Blame x SXSWFollower2007	-0.0000***	-0.0000***	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Merit x SXSWFollower2007	2.4824***	1.0283***	2.4824***	1.0283***
	(0.1796)	(0.1290)	(0.1796)	(0.1290)
Blame x SXSWFollower2006	0.0000***	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Merit x SXSWFollower2006	-0.2744	-0.5627**	-0.2744	-0.5627**
	(0.3938)	(0.2451)	(0.3938)	(0.2451)
Politician FE	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
Extended Controls		✓		✓
Sentiment Control			✓	✓
Observations	245,532	245,453	245,532	245,453
Clusters	3,108	3,107	3,108	3,107
F statistic	0.93	0.97	0.93	0.97
Partial F statistic Blame x User	190.99	63.52	190.95	63.52
Partial F statistic Merit x User	190.99	63.60	190.95	63.60

Notes: The table presents the first-stage regressions related to the 2SLS estimation of Equation 5. In Panel A) the outcome is the monthly share of blame tweets posted by all politicians interacted with the log+1 number of Twitter users in the county. In Panel B) the outcome is the monthly share of merit tweets posted by all politicians interacted with the log+1 number of Twitter users in the county. In each panel, the outcome is regressed on the monthly share of blame and merit tweets posted by all politicians interacted with the log+1 number of SXSW followers in the county who joined in 2007. The table also includes the estimates for the same tweet shares interacted with the log+1 number of SXSW followers in the county who joined in 2006. Standard errors in parentheses clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A21: Survey Questions

Variable	Question Text	Answers	Coding
Affective polarization	Standard thermometer question.	0 – 100	Used as it is
Trust in politicians	If a member of Congress were offered a bribe to influence the awarding of a government contract, do you think that the member of Congress would accept or refuse the bribe?	Extremely likely to refuse Likely to refuse Equally likely to refuse or accept Likely to accept Extremely likely to accept	1 if extremely likely or likely to refuse 0 otherwise
Government responsiveness	If you were to complain about the poor quality of a public service, how likely or unlikely is it that the problem would be easily resolved?	Extremely unlikely Unlikely Equally likely or unlikely Likely Extremely likely	1 if extremely likely or likely 0 otherwise

Notes: The table presents details about the survey questions used for the analysis in Section 6.

Table A22: Attitudes

	Affective Polarization	Trust in Politicians	Government Responsiveness
Blame	0.112**	-0.138***	-0.065*
	(0.045)	(0.050)	(0.034)
Merit	0.020	0.067	0.073*
	(0.050)	(0.043)	(0.040)
Week FE	✓	✓	✓
Observations	2,099	2,099	2,099
Clusters	51	51	51

Notes: The table presents the estimates of Equation 6. In the first column the outcome is the standardized average level of affective polarization among respondents. In the second column the outcome is the standardized average level of trust in politicians among respondents. In the third column the outcome is the standardized average level of perceived government responsiveness among respondents. Standard errors in parentheses clustered at the state level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

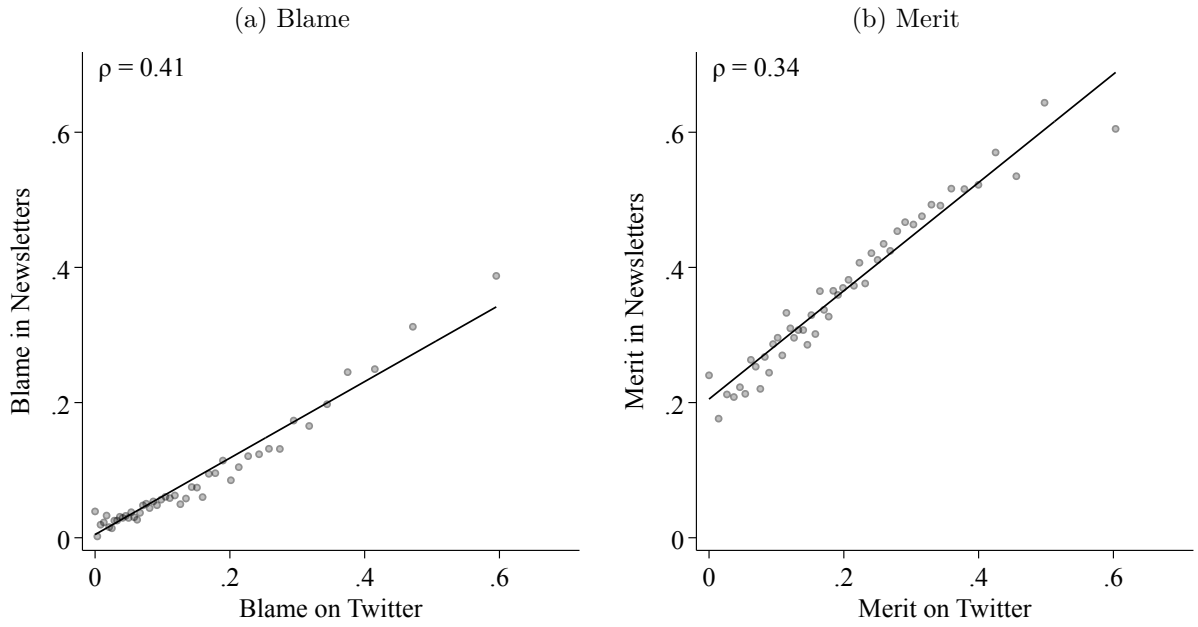
B Additional Results

B.1 Newsletters

A limitation of our analysis is its exclusive focus on one platform, Twitter. Two arguments support this choice. First, Twitter's focus on the sharing of opinions has made it a standard for studies at the intersection of media economics and political economy (Barberá et al., 2019; Demszky et al., 2019; Halberstam and Knight, 2016, e.g.,). Second, by concentrating on communication from politicians to voters, our study naturally limits the range of platforms amenable to analysis – other potential sources include presidential speeches, political ads, manifestos, and political newsletters. As a first step towards external validity, we adapt our classifier to newsletter data coming from Cormack (2025). This data source presents some conceptual differences from the Twitter data. First, it represents a more tailored communication channel between a politician and their constituency, which means that dialogue is not restricted by platform policies or indirectly influenced by algorithms. Second, for these reasons, this is potentially a more institutional channel of communication. To apply our classifier, we analyze the newsletter corpus at the sentence-level, so that text-units have comparable length to tweets. Then, for each newsletter piece, we mimic our tweet classification by computing the share of merit, blame, and none sentences.

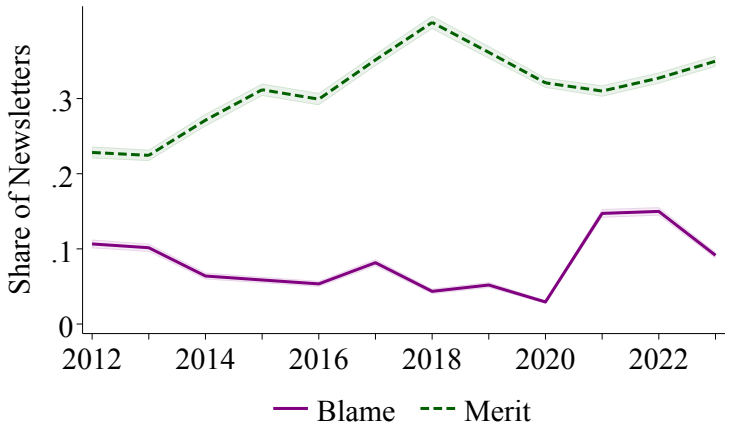
First, we compare politicians' reliance on blame and merit on Twitter and in newsletters. Figure B1 shows that, for each politician in each quarter, the share of blame (merit) tweets is highly and positively correlated with the share of blame (merit) sentences in their newsletter communications. This correlation, despite the contextual differences mentioned above, strongly suggests that the dimension we are capturing in our analysis is not strictly related to Twitter. To further investigate this point, Figure B2 plots the evolution of blame and merit in newsletters. One noticeable difference is that the merit share starts and remains considerably higher than the blame share; this is intuitive given the more institutional nature of these messages. However, one can appreciate a similar qualitative pattern with an increase over time on both dimensions: in 2022 the share of blame and merit sentences is 50 percent higher than in 2012.

Figure B1: Blame and Merit: Twitter vs. Newsletters



Notes: In Panel (a) we plot on the y-axis the share of blame sentences for each politician in each quarter, while on the x-axis the share of blame tweets for the same politician in the same quarter. Panel (b) does the same for the share of merit sentences. Observations are split in 50 bins with the `binscatter` command.

Figure B2: Supply of Blame and Merit Newsletters over Time



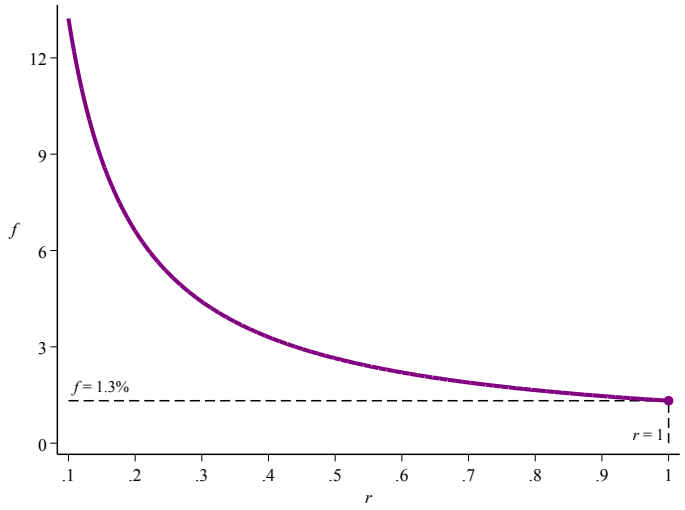
Notes: The figure presents the yearly share of newsletter sentences classified as blame and merit. Shaded areas represent 95 percent confidence intervals.

B.2 Persuasion Rate

To interpret the magnitude of our estimates, we calculate persuasion rates following DellaVigna and Gentzkow (2010) as $f = \frac{y_c - y_t}{r \times e_t - e_c} \times \frac{1}{1-y_0}$. In the expression $y_c - y_t$ is the difference in the share of population donating between treated and untreated, y_0 is the share that would donate without the treatment, $e_t - e_c$ is the difference in exposure between the treated and untreated, and r is a reach parameter. In our setting, we can write $y_t - y_c = \bar{\beta} \times y_c = \beta \times \overline{\text{Users}} \times y_c$, where β is the semi-elasticity of log number of donors

to a specific rhetorical style and $\overline{\text{Users}}$ is the average county log Twitter penetration. Since the estimated β for merit is close to 0 and statistically insignificant, we focus on blame. As standard in the literature, we assume $y_0 = y_c$, and we set $y_c = 0.16$ following Boken et al. (2023). The $r \times e_t - e_c$ term deserves more discussion. While e_c is mechanically equal to 0, e_t represents the share of the average county that is on Twitter (again, 0.32 in Boken et al. (2023)). For our purposes, however, the relevant exposure metric is the share of the average county that could have been exposed to the rhetorical style. For instance, since they focus on virality, Boken et al. (2023) estimate assumes that a viral tweet will be seen by everybody who is on Twitter, effectively setting $r = 1$. This is the same assumption that leads to the 1.3% persuasion rate reported in the main text, which is a lower bound. However, there are reasons to believe that r could be lower. For instance, Wojcieszak et al. (2022) find that 40% of American Twitter users follow at least one politician. Assuming that users see only the posts of accounts they follow, and they see all of them, this would lead to $r = 0.4$ and thus $f = 3.3\%$. Hence, to be as transparent as possible, we report estimates of how f changes as the assumed reach r changes in Figure B3. Reassuringly, we see that the implied magnitude for f remains meaningfully interpretable as r decreases, with a maximum of $f \approx 13\%$ at $r = 0.1$, which would be smaller than the Fox News effect documented in DellaVigna and Kaplan (2007).

Figure B3: Blame's Persuasion Rate and Reach



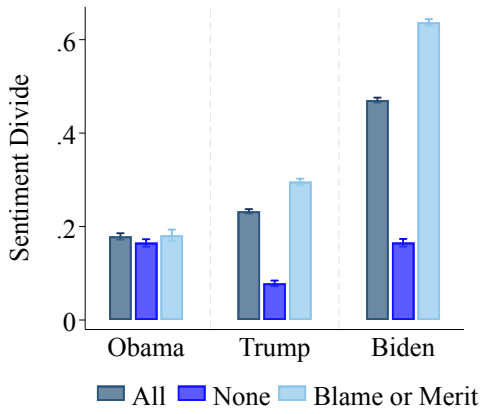
Notes: The figure plots the calculated value of f for given values of $r \in [0.1, 1]$.

B.3 Elite Polarization

We undertake several analyses to address potential robustness concerns of our findings. One may worry that blame and merit tweets, by construction, reference political actors and are therefore more likely to carry a partisan slant, whereas non-causal tweets may not target specific political actors. To alleviate this concern, we show in Figure B4 that

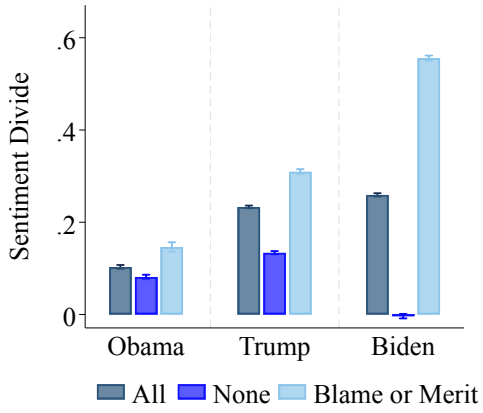
results are unchanged if we restrict only to those tweets that we can clearly identify as targeting either Democrats or Republicans. Furthermore, our findings persist along the intensive margin – how strongly positive or negative a tweet is. Figure B5 replicates the analysis excluding neutral tweets. The results remain unchanged, presenting the same pattern, and suggesting that causal rhetoric is where politicians polarize increasingly over time.

Figure B4: Elite Polarization Restricting to Targeted Tweets



Notes: The figure difference in average standardized sentiment between tweets posted by members of the ruling party and those posted by members of the opposition, separately for each presidency. We compute this difference across three subsamples: all tweets, tweets that are neither blame nor merit, and tweets that are either blame or merit. In all three subsamples, we only include tweets that have been identified to target a political actor. In both panels, bars represent 95 percent confidence intervals.

Figure B5: Elite Polarization Excluding Neutral Tweets



Notes: The figure presents the difference in average standardized sentiment between tweets posted by members of the ruling party and those posted by members of the opposition, separately for each presidency. We compute this difference across three subsamples: all tweets, tweets that are neither blame nor merit, and tweets that are either blame or merit. In all three subsamples, we exclude tweets that are classified with a neutral sentiment, that is, including only tweets classified as positive or as negative. In both panels, bars represent 95 percent confidence intervals.

B.4 Virality

This Appendix provides more details regarding the relationship between rhetorical style and virality discussed in Section 4. We first discuss differences in the availability of engagement measures. Then, we present evidence that blame gets more retweets and is more associated with virality, while merit does not. Finally, we show that this pattern has intensified over time.

We observe engagement data only for a subsample of the main dataset used throughout the paper. In particular, retweet information is available until March 2020, covering approximately half of all tweets in the original sample for this time period. Table B1 compares the full dataset and the engagement subsample over this period. Concerning differences at the politician level, tweets with engagement data tend to come from politicians who are, on average, one year older and 2 percentage points more likely to be Republican. However, differences in demographic and ideological characteristics are uniformly small in economic terms. Then, focusing on tweet-level again, differences are limited in magnitude: tweets with engagement data display a slightly higher share of merit tweets – about 4 percentage points more – while other dimensions remain closely aligned. Taken together, these comparisons suggest that the engagement subsample is broadly representative of the full dataset, alleviating concerns about sample selection bias in the analysis that follows.

To show how blame and merit correlate with engagement, we estimate the following specification:

$$y_{ipt} = \beta_1 \text{Blame}_i + \beta_2 \text{Merit}_i + \delta_1 \text{Sentiment}_i + \lambda_p + \mu_t. \tag{7}$$

Here y_{ipt} is the standardized retweet count of tweet i posted by politician i about topic t; Blame_i and Merit_i are binary indicators taking the value 1 if the tweet is blame or merit, respectively; λ_p and μ_t denote politician and topic fixed effects, respectively. Results, shown in Figure B6, indicate that blame tweets receive, on average, 0.2 standard deviations more retweets than tweets with no causal rhetoric. By contrast, merit tweets do not get significantly more retweets than non-causal tweets.

To assess how the relationship between rhetorical style and engagement varies across the distribution of retweets, we re-estimate the specification from Equation 7 ten times, each time using as the dependent variable an indicator for whether a tweet falls into each of the ten deciles of the retweet distribution. That is, each regression estimates the association between blame and merit and the probability of falling into each specific decile. Figure B7 presents the results. Blame tweets are significantly more likely to appear in the right tail of the distribution – particularly in the top decile – confirming their association with virality. By contrast, merit tweets are modestly more likely to appear just above the median, but show no higher probability of reaching the top deciles.

Finally, we examine how the relationship between rhetorical style and virality has evolved over time. To do so, we estimate the following specification:

$$y_{iptj} = \alpha + \sum_{k=2013}^{2020} \beta_1^k \left(\text{Blame}_i \times D_{ij}^k\right) + \sum_{k=2013}^{2020} \beta_2^k \left(\text{Merit}_i \times D_{ij}^k\right) + \lambda_p + \mu_t + \eta_j + \sum_{k=2013}^{2020} \delta_1^k \left(\text{Sentiment}_i \times D_{ij}^k\right) + \varepsilon_{iptj}. \quad (8)$$

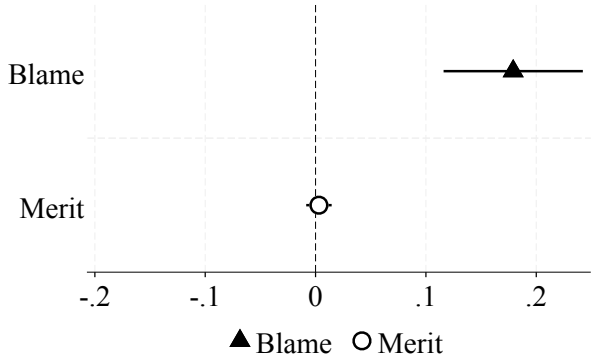
Here y_{iptj} is the standardized retweet counts of tweet i posted by politician i about topic t in year j; Blame_i and Merit_i are binary indicators taking the value 1 if the tweet is blame or merit, respectively; D_j^k are year indicators taking the value 1 if tweet i is posted during year k; λ_p, μ_t, and η_j denote politician, topic, and year fixed effects, respectively. Finally, we also include the sentiment of the tweet to isolate the role of blame and merit from pure sentiment. This allows us to compare the relative virality of blame and merit tweets over time, relative to the baseline year of 2012. Figure B8 presents the results. Before 2016, neither blame nor merit tweets received systematically more engagement than non-causal tweets. Starting in 2017, however, blame tweets became significantly more viral, reaching an average effect size of approximately 0.4 standard deviations by 2020. In contrast, the relative engagement of merit tweets remained flat or slightly declined over the same period.

Table B1: Differences Between With/Without Engagement Datasets

	All Tweets until 03/2020	Tweets with Retweet Data	Difference	p-value
Female	0.275 (0.000)	0.276 (0.000)	0.001	0.196
Age	57.711 (0.007)	58.665 (0.010)	0.955	0.000
Black	0.076 (0.000)	0.074 (0.000)	-0.002	0.000
Bachelor	0.329 (0.000)	0.336 (0.000)	0.007	0.000
Master or Higher	0.626 (0.000)	0.617 (0.000)	-0.009	0.000
Republican	0.431 (0.000)	0.450 (0.000)	0.019	0.000
|Nominate Score|	0.434 (0.000)	0.429 (0.000)	-0.005	0.000
Share of Blame Tweets	0.144 (0.000)	0.148 (0.000)	0.004	0.000
Share of Merit Tweets	0.168 (0.000)	0.206 (0.000)	0.038	0.000
Share of None Tweets	0.687 (0.000)	0.646 (0.000)	-0.042	0.000
Share of Tweets about Economy	0.150 (0.000)	0.151 (0.000)	0.001	0.064
Share of Tweets about Environment	0.051 (0.000)	0.054 (0.000)	0.004	0.000
Share of Tweets about Gender	0.070 (0.000)	0.064 (0.000)	-0.006	0.000
Share of Tweets about Gun Control	0.032 (0.000)	0.032 (0.000)	-0.000	0.326
Share of Tweets about Healthcare	0.118 (0.000)	0.122 (0.000)	0.003	0.000
Share of Tweets about Immigration	0.055 (0.000)	0.059 (0.000)	0.003	0.000
Share of Tweets about Police	0.027 (0.000)	0.027 (0.000)	0.000	0.042
Share of Tweets about Racial Relations	0.064 (0.000)	0.059 (0.000)	-0.005	0.000
Share of Tweets about Other Topics	0.433 0.000	0.432 0.000	-0.001	0.184
Observations	2322957	1178861		

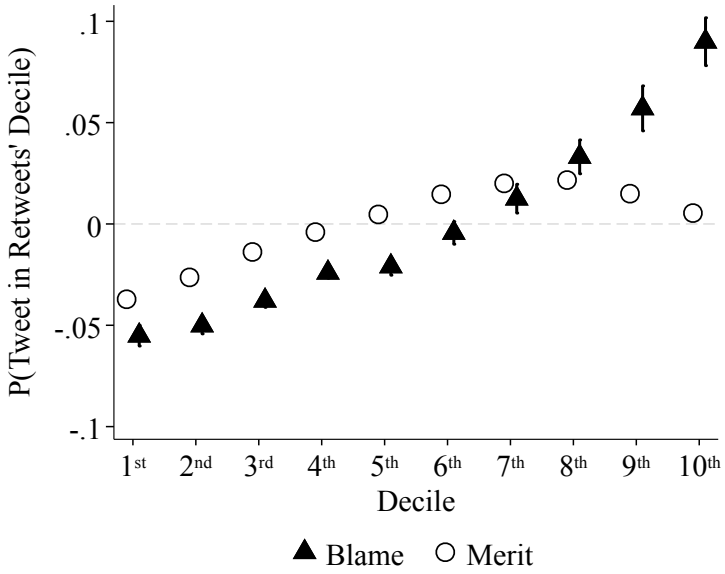
Notes: Standard errors in parentheses.

Figure B6: Blame, Merit, and Retweets



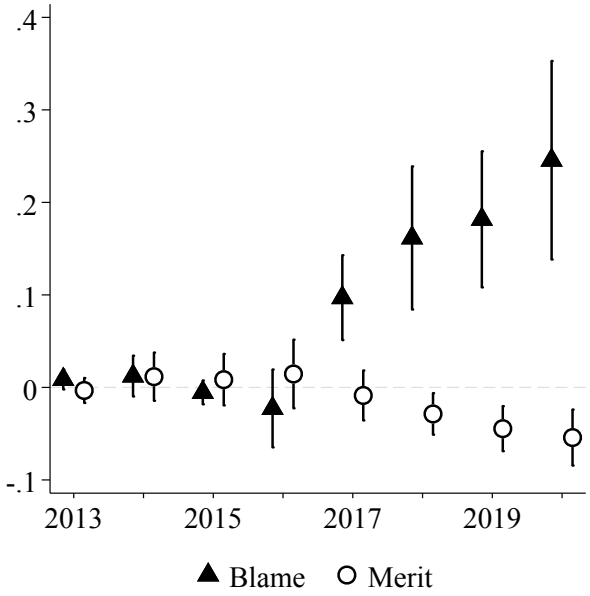
Notes: Bars represent 95 percent confidence intervals, errors clustered at the politician level.

Figure B7: Blame, Merit, and Retweets' Distribution



Notes: Bars represent 95 percent confidence intervals, errors clustered at the politician level.

Figure B8: Blame, Merit, and Engagement over Time



Notes: Shaded areas represent 95 percent confidence intervals, errors clustered at the politician level.

B.5 Platform Changes

As discussed in Section 3, we document a substantial increase in the supply of blame and merit tweets by politicians, particularly beginning in 2017. A potential concern is that this shift may have been driven by platform-level policy changes implemented by Twitter during our sample period. Two major changes are worth considering. First, on February 10, 2016, Twitter introduced algorithmic feed curation, replacing the strictly chronological ordering of tweets. Second, on November 7, 2017, Twitter doubled the character limit from 140 to 280 characters. For both events, we find no evidence of a discontinuity in the production of blame and merit tweets, making it unlikely that either change is responsible for the observed rise in their usage.

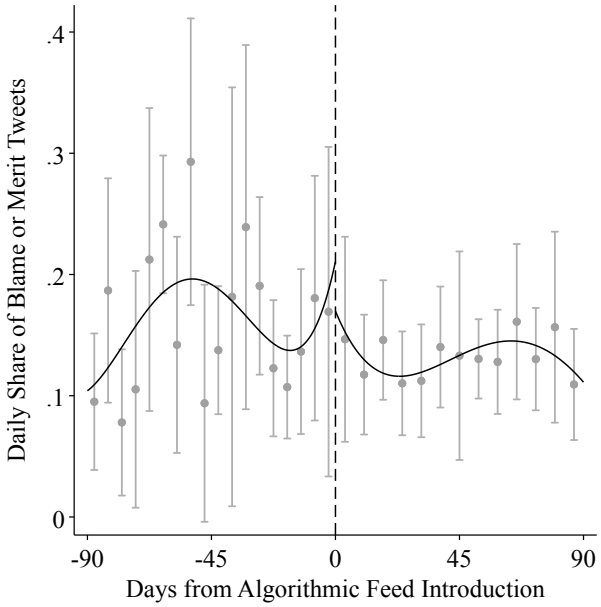
We begin by considering the introduction of the algorithmic feed on February 10, 2016. First, it appears unlikely that this change drove an increase in the supply of blame and merit tweets, as Figure 2 shows no notable rise in their use during 2016. To provide more rigorous evidence, we estimate a regression discontinuity design centered on the policy change, using the daily share of tweets classified as blame or merit as the outcome. The analysis covers a symmetric window of 90 days around February 10. As shown in Figure B9, there is no evidence of a discontinuity at the threshold. The estimated jump is -0.035 and statistically insignificant ($p = 0.529$). These results confirm that the introduction of the algorithmic feed does not appear to have contributed to the rise in causal rhetoric.

We examine the character-limit change in greater detail. In Figure B10, we plot the

distribution of tweet lengths in characters before November 7, 2017. The figure shows clear bunching near the upper limit, particularly in the 100-120 character range, indicating that many tweets were close to exhausting the 140-character constraint. However, this pattern is similar across rhetorical categories: blame, merit, and none tweets all exhibit similar length distributions. This suggests that the original character limit was not disproportionately constraining the production of blame or merit tweets. If blame and merit tweets had been uniquely limited by the character cap, we would expect to observe stronger bunching near the limit for those categories. Instead, their length distribution closely mirrors that of none tweets. It is therefore unlikely that the doubling of the character limit in 2017 "freed up" the production of blame or merit tweets, or that it played a meaningful role in the subsequent increase in their supply.

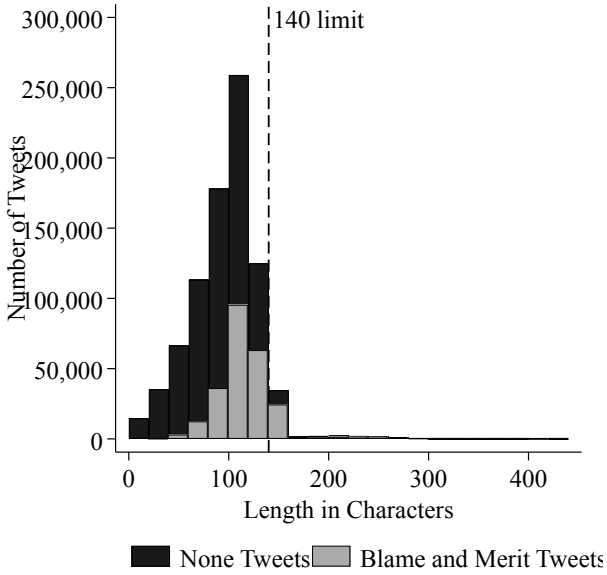
To provide more formal evidence, we carry out a regression discontinuity analysis around the date of the character limit expansion, using the daily share of tweets classified as blame or merit as the outcome. The analysis focuses on a symmetric window of 90 days around November 7, 2017. As shown in Figure B11, we find no evidence of a discontinuity at the cutoff. The estimated jump is 0.056 and statistically insignificant, with a p-value of 0.532. This result reinforces the conclusion that the character limit expansion did not play a pivotal role in the rise of causal rhetoric on the platform.

Figure B9: Regression Discontinuity Plot around Algorithmic Feed Introduction



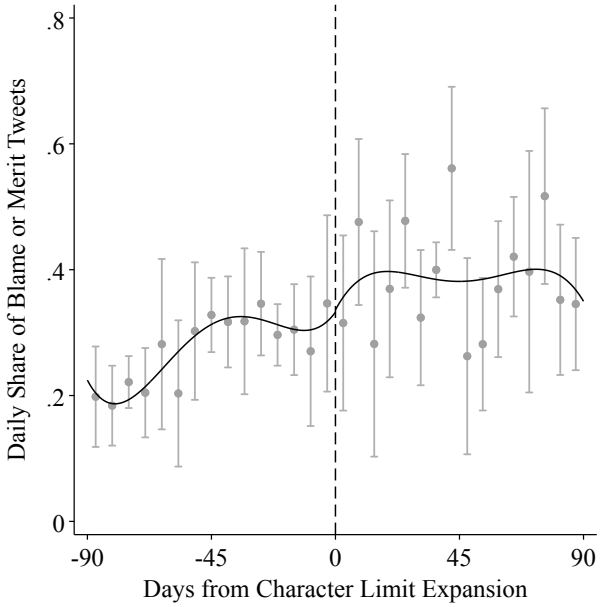
Notes: The figure presents the regression discontinuity plot around the introduction of the algorithmic feed produced with the `rdrobust` package by Calonico et al. (2017).

Figure B10: Distribution of Tweet Lengths before Character Limit Expansion



Notes: The figure presents the distribution of tweet lengths until November 7, 2017.

Figure B11: Regression Discontinuity Plot around Character Limit Expansion



Notes: The figure presents the regression discontinuity plot around the character limit expansion produced with rdrobust package by Calonico et al. (2017).