

Blameocracy: Causal Rhetoric in Politics*

Francesco Bilotta

Alberto Binetti

Giacomo Manferdini

Job Market Paper

August 20, 2026

[Click here for the latest version](#)

Abstract

This paper studies the supply and returns of causal rhetoric in politics: politicians' language linking political actions to outcomes through blame and credit. We fine-tune a language model to measure its use in 4.2 million congressional tweets spanning 2012–2023. Causal rhetoric is central to politicians' communication: it accounts for more than one third of tweets, is the language of political disagreement, and rises over time. Political returns are concentrated in blame. Combining campaign contributions with quasi-exogenous geographic variation in Twitter adoption, we find that exposure to blame has the largest effects on small donations. Using voting intentions from survey data and an event-study design around blame-only and credit-only days, we find that politicians' support rises in the week after blame-only days. Additional evidence suggests causal rhetoric, and especially blame, can turn complex economic mechanisms into underqualified yet plausible, politically convenient claims.

*Francesco Bilotta: Bocconi University, francesco.bilotta2@phd.unibocconi.it; Alberto Binetti: Princeton University, abinetti@princeton.edu; Giacomo Manferdini: Bocconi University, giacomomanferdini@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, Felix Chopra, Sarah Eichmeyer, Nicola Gennaioli, Gloria Gennaro, Dirk Hovy, Rafael Jiménez-Durán, Tania Lombrozo, Massimo Morelli, Jaime Marques Pereira, Edoardo Teso, Egon Tripodi, Carlo Schwarz, and Ekaterina Zhuravskaya for their helpful comments and suggestions, as well as participants to the 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, 1st Berlin Micro Theory and Behavioral Economics PhD Conference, 10th Monash-Paris-Warwick-Zurich-CEPR Text-as-Data Workshop, 1st CEPR Future of Democracy Conference at WZB, Ludwig Erhard ifo Research Workshop, 18th UniTo-CCA PhD Workshop in Economics, XXIII RIDGE-LACEA Political Economy Workshop, 9th Economics of Media Bias Workshop, and 2nd Berlin PhD Conference in Economics.

“To err is human. To blame someone else is politics.”
Hubert H. Humphrey, Vice President of the United States (1965–1969)

1 Introduction

Democratic politics asks voters to make a difficult causal judgment. They must form beliefs about how political actions map into outcomes: who is responsible for what has happened, and what consequences political actors are likely to produce in the future (Schmitter and Karl, 1991). Yet this mapping is rarely transparent. Policies operate in complex environments, outcomes have many causes, and responsibility is often shared across parties, institutions, and levels of government. Voters therefore face the problem of reconstructing a mapping between political actions and outcomes that is politically consequential but hard to infer (Achen and Bartels, 2016).

This difficulty creates a rhetorical margin: politicians can shape voters’ understanding of this mapping. We call this margin *causal rhetoric*: language that links the actual or potential action of a political actor to an outcome. When the outcome is negative, politicians assign *blame*; when it is positive, they claim *credit*. In both cases, causal rhetoric supplies voters with an account of how political actions and outcomes are connected, bringing a well-documented persuasive force to a central problem of democratic politics. This force comes from the way causal accounts shape the information people seek, remember, and act on (Andre et al., 2026; Bursztyn et al., 2023; Graeber et al., 2026; Hüning et al., 2022).¹ In politics, these accounts are a first step toward accountability: they tell voters which outcomes to link to which political actions. Yet we know little about how politicians supply causal rhetoric or the political returns they gain from using it.

This paper studies these questions in the context of U.S. congressional communication on Twitter. We fine-tune a language model to build a content-based measure of causal rhetoric, classifying tweets by whether they assign blame or claim credit. We then combine this measure with data on campaign contributions and voting intentions to document the supply of causal rhetoric and estimate its political returns. The results show that causal rhetoric is a central part of politicians’ communication and that its political returns are concentrated in blame. Finally, examining economic claims suggests that causal rhetoric, and especially blame, can turn complex economic mechanisms into underqualified yet plausible, politically convenient claims.

The measurement starts from a corpus of 4.2 million tweets posted by Democratic

¹This evidence is consistent with a broader literature in cognitive psychology emphasizing that causal thinking is central to human reasoning (e.g., Chater and Loewenstein, 2016; Lombrozo and Vasiljeva, 2017; Sloman and Lagnado, 2015).

and Republican members of Congress between 2012 and 2023. We hand-label a training set of tweets and use these labels to fine-tune a RoBERTa-large model pre-trained on Twitter data (Loureiro et al., 2022), classifying each tweet as blame, credit, or neither. The resulting measure performs well against hand-coded validation data and aligns closely with independent external classifications. Additional validation shows that blame and credit exhibit the linguistic patterns expected of causal attribution and are not simply a relabeling of previously studied dimensions of political language. This yields a scalable measure of causal rhetoric in politicians’ communication on social media.

We first establish that causal rhetoric is a central part of politicians’ communication on social media. More than one third of tweets contain causal rhetoric: 16 percent assign blame and 20 percent claim credit. Its use is more concentrated in policy domains than in sociocultural domains, consistent with politicians using it where political actions can be more naturally linked to outcomes. Causal rhetoric is also the language of political disagreement. The sentiment gap between the ruling party and the opposition is much larger in tweets that assign blame or claim credit than in other tweets, indicating that politicians’ evaluative disagreement is especially concentrated in causal claims. Finally, the aggregate share of blame and credit rises over time, from roughly one fifth of congressional tweets at the beginning of the sample to nearly one half by the end, with both components increasing.

This prevalence motivates the paper’s main returns question: does causal rhetoric generate political support? Returns reveal whether this widespread rhetorical input is politically rewarded, rather than merely supplied, and whether it affects citizens’ political behavior. We find that these returns are concentrated in blame. Exposure to blame generates sizable and consistent gains in support, while credit produces no comparable aggregate response. We establish this pattern using two complementary sources of evidence: campaign contributions and survey data on voting intentions.

Our first source of evidence is campaign contributions. We link our rhetoric measure to more than 40 million small donations, defined as individual contributions below \$1,000 (Boken et al., 2023; Bonica, 2024; Petrova et al., 2021). To estimate returns, we use a cross-county shift-share-style exposure design (Borusyak et al., 2025; Goldsmith-Pinkham et al., 2020) that combines politician-month variation in blame and credit with quasi-exogenous geographic variation in Twitter adoption induced by the platform’s early diffusion after the South by Southwest festival (Fujiwara et al., 2024; Müller and Schwarz, 2023). The identifying comparison asks whether counties with greater quasi-exogenous Twitter exposure donate more to the same politician in months when that politician uses more causal rhetoric, net of politician-month, politician-county, and county-month fixed effects. Because the design relies on quasi-

exogenous differences in exposure rather than on short-run timing around tweets, it does not require politicians to choose blame or credit at random.

The donation estimates show asymmetric returns, larger for blame. A one-standard-deviation increase in the monthly share of blame tweets raises small-donation revenue by roughly 2.6 percent in a county with average Twitter adoption, while the corresponding effect of credit is about 0.5 percent. These aggregate effects operate through different margins of support. Blame expands the donor base, increasing the number of donors by about 1.6 percent, whereas credit has essentially no effect on donor counts and its small return is concentrated in higher average giving. The returns also differ by donor ideology, suggesting that blame and credit activate different constituencies within the donor base: ideologically extreme donors respond positively to blame and negatively to credit, while moderate donors display the opposite pattern.

We complement the donation evidence with survey data on voting intentions in the 2020 House elections. Using 92,211 responses from Nationscape ([Tausanovitch and Vavreck, 2021](#)), we link respondents to the Twitter communication of the sitting representative in their congressional district. This setting lets us measure political support directly as intended vote choice, extend the analysis beyond donors, and use a short-run timing comparison. We implement an event-study design around days when a politician's Twitter communication consists entirely of one rhetorical style. The design asks whether support for the same politician changes after blame-only or credit-only days, net of district-month, date, partisan-alignment, and respondent-demographic fixed effects.

The survey estimates show an asymmetric response in intended vote choice. Support for the politician rises by 1 percentage point in the week after blame-only days. The corresponding estimate after credit-only days is smaller and imprecisely estimated. A dynamic version of the design shows flat pre-trends before blame-only days, followed by a short-run increase in support that fades within two weeks. The response is concentrated among respondents who report using social media for political news: among these respondents, support rises by 1.5 percentage points after blame-only days, compared with a smaller and imprecise estimate among other respondents. The survey evidence therefore points to the same asymmetry as the donation evidence and reinforces the social-media exposure channel.

Retweet activity points to amplification as one reason blame generates larger returns. In the subsample with non-missing retweet data, blame accounts for 15 percent of tweets but nearly 40 percent of all retweets, and blame tweets receive almost four times as many retweets as other tweets. More structured comparisons yield the same conclusion: blame travels farther than credit or non-causal rhetoric. This greater reach points to amplification as a channel consistent with blame's larger political returns.

The returns evidence raises a further interpretive question: what kind of claims are contained in the rhetoric that politicians widely use and that receives political returns, especially in blame? We study this question in the economic domain, where causal claims are common and can be compared to standard macroeconomic reasoning.

The evidence points to claims that are underqualified yet plausible, and politically convenient in how they assign blame and credit. Using an AI-assisted coding procedure to evaluate each claim's support under standard macroeconomic reasoning, we find that most economic causal claims require qualifications that the tweet does not provide, while outright unsupported claims are rare. This underqualification is especially pronounced for blame. We then use economic bill introductions to examine whether the use of these claims lines up with observable political reality and whether it does so in a politically convenient direction. Economic blame rises around opposing-party legislative action, while economic credit rises around own-party legislative action. The evidence therefore points to a subtle margin: causal rhetoric, and especially blame, can turn complex economic mechanisms into underqualified yet plausible, politically convenient claims.

Taken together, the evidence shows that causal rhetoric is a major input into democratic politics. Politicians widely supply voters with claims about how political actions map into outcomes, and these claims generate political returns when they assign blame. In the economic domain, these claims are not usually outlandish, but neither are they usually grounded as stated: they are often underqualified, plausible, and politically convenient in how they assign blame and credit. This pattern is sharpest for blame, where political returns are concentrated.

Related Literature We contribute to the literature on accountability in political economy by studying how politicians seek to shape the attributions on which accountability rests. A large body of work studies how accountability shapes politicians' behavior, showing, for example, how term limits and press coverage discipline elected officials (Ashworth, 2012; Besley, 2006; Besley and Case, 1995; Canes-Wrone et al., 2001; Ferraz and Finan, 2011; Maskin and Tirole, 2004; Snyder Jr and Strömberg, 2010). A related strand studies how voters use realized outcomes to evaluate incumbents, emphasizing both the logic and the limits of retrospective voting (Achen and Bartels, 2016; Ferraz and Finan, 2008; Healy and Malhotra, 2013). We study an upstream margin in this process: politicians' use of rhetoric that links political actions to outcomes. We document the prevalence of this rhetoric and estimate its political returns.

This paper adds to the political-economy literature on social media by identifying a specific content channel through which online communication affects political support. A growing body of work shows that social media affects political support

and political behavior (Fujiwara et al., 2024; Müller and Schwarz, 2023; Rotesi, 2019; Zhuravskaya et al., 2020). Closest to our paper is work on Twitter and campaign finance: Petrova et al. (2021) show that politicians benefit from adopting Twitter, especially entrants and candidates in high-adoption states, while Boken et al. (2023) document a fundraising premium when politicians go viral on Twitter. We shift attention from platform adoption, diffusion, and virality to the content of politicians' messages. Rather than treating social-media exposure as a single reduced-form treatment, we identify blame and credit as rhetorical inputs through which politicians' online communication generates political support.

We also contribute to the literature on narratives in political economy by measuring a broad reduced-form feature of political narratives: whether politicians connect political actions to outcomes. Existing work measuring narratives in observational data has mainly focused on specific topics, such as climate change or stock-market crashes (Gehring and Grigoletto, 2025; Goetzmann et al., 2024; Macaulay and Song, 2024). We take a more agnostic approach, classifying whether politicians' language supplies a causal link between political actors and outcomes, regardless of the topic under discussion. This measure connects to Bilotta and Manferdini (2026), where politicians compete by shifting blame and claiming credit when voters cannot fully recover the relationship between policies and outcomes. Our evidence documents the prevalence and political returns of this rhetoric, and shows that these claims, despite being under-qualified, remain plausible by staying minimally consistent with observable political reality.²

Finally, we contribute to the literature on political rhetoric by isolating the causal margin of political language: whether politicians link political actions to outcomes. A long-standing literature on negative campaigning studies whether campaigns attack opponents or promote the candidate (Lau and Rovner, 2009). We focus instead on blame as a specific form of negative rhetoric: language that connects political actors or their actions to bad outcomes. This excludes other negative content, such as personal attacks, insults, or moral condemnation, when it does not contain a causal claim. We also complement recent work on expressive dimensions of political language, such as emotionality, emotions, and moral values (Algan et al., 2026; Enke, 2020; Gennaro and Ash, 2022). We instead identify causal rhetoric as a distinct dimension of political discourse.

The rest of the paper proceeds as follows. Section 2 describes the data. Section 3 defines causal rhetoric through blame and credit, explains how we measure it, and

²More broadly, the idea that narratives should remain at least minimally consistent with evidence is a feature shared by several other theoretical models of narratives (e.g., Eliaz et al., 2025; Eliaz and Rubinstein, 2025; Eliaz and Spiegler, 2020).

validates the resulting classification. Section 4 documents the prevalence, partisan-conflict patterns, and trends of causal rhetoric in congressional communication. Section 5 estimates its political returns using campaign-contribution and voting-intention data. Section 6 interprets economic causal rhetoric by combining a support-as-stated content analysis with evidence from economic bill introductions. Finally, Section 7 concludes.

2 Data

Our analysis combines several datasets on political communication, campaign finance, survey support, and legislative activity.

Tweets Our main dataset contains approximately 4.2 million tweets posted by Democratic and Republican members of the U.S. Congress between 2012 and 2023. We combine data from the [Tweets of Congress](#) project and [Bellodi et al. \(2025\)](#). The resulting dataset covers the universe of tweets posted by Representatives from January 3, 2013, to July 11, 2023, and by Senators from June 21, 2017, to July 11, 2023. It also covers all tweets posted over the full 2012 calendar year by Representatives running for re-election that year. The final sample includes 4,198,455 tweets from 900 unique politicians and 1,789 unique Twitter accounts.³ For a subset of original tweets, we also observe retweet counts. We further enrich these data with demographic and ideological information from VoteView and Wikipedia, including age, gender, race, education, and NOMINATE scores. Table A1 reports descriptive statistics at both the tweet and politician levels. Republicans account for a slight majority of politicians in the sample, but most tweets are posted by Democrats, who—as also documented by [Fujiwara et al. \(2024\)](#)—are more active on Twitter. Most politicians in the dataset hold at least a bachelor’s degree, the average age is just under 60, and about one quarter are women.

Donations We collect campaign contribution data from the Database on Ideology, Money in Politics, and Elections compiled by [Bonica \(2024\)](#). This dataset includes more than 850 million donations made by individuals and organizations to candidates in local, state, and federal elections between 1979 and 2024. We restrict the sample to individual contributions directed to candidates for the U.S. House of Representatives and U.S. Senate in the 2012–2024 election cycles. Following the literature, we define small donations as contributions below \$1,000 ([Boken et al., 2023](#); [Petrova et al., 2021](#)).

³We exclude retweets because they do not represent original content. We retain quoted tweets because they also contain user-generated text.

In the analysis, we link these donations to politicians in our Twitter data, so the effective sample period is determined by the overlap between donation records and observed tweeting behavior and therefore runs through 2023. The resulting matched sample contains 40,087,963 donations to 733 politicians.

Voting Intentions We use Nationscape survey data for the 2020 House elections. Nationscape is a repeated cross-section that records respondents’ congressional district, intended House vote choice, party identification, demographic characteristics, and recent political-news sources (Tausanovitch and Vavreck, 2021). To ensure that the relevant general-election choice is defined, we restrict the sample to respondents interviewed after the relevant congressional primary in their state, using official primary dates from [Federal Election Commission](#).⁴ We keep respondents interviewed through October 31, 2020, just before the general election, and restrict attention to districts in which the sitting representative is running for re-election, since our tweet data cover sitting legislators. This leaves us with 92,211 survey responses.

Bills We also construct a dataset of congressional bills using official records from the U.S. Government Publishing Office, accessed through the GovInfo bulk data repository. We scrape structured metadata for all bills introduced in the U.S. House and Senate from the 112th Congress (2011–2013) through the 118th Congress (2023–2025).⁵ To focus on politically salient legislation, we restrict attention to bills sponsored by Democratic or Republican legislators that received at least one roll-call vote. Because the analysis compares economy-related legislative activity across chambers, we further restrict the sample to the period in which Senators are observed in our Twitter data and to bills whose metadata identify them as belonging to an economic policy area.⁶ The resulting economic-bill sample contains 327 bills.

America’s Political Pulse We use data from America’s Political Pulse, a project led by the Polarization Research Lab that also tracks elite political rhetoric (Westwood and Lelkes, 2024). The project classifies congressional communications—including speeches, newsletters, tweets, press releases, and public statements—into rhetorical

⁴In states where a runoff can determine the nominee, we use the runoff date instead. Louisiana is treated separately because it does not hold a separate congressional primary; for Louisiana, we use the congressional ballot-access filing deadline.

⁵Specifically, we include House bills, House joint resolutions, House concurrent resolutions, and House simple resolutions, as well as their Senate counterparts.

⁶We identify economic bills using the policy-area field included in the GovInfo metadata. We keep bills in the following policy areas: Agriculture and Food, Commerce, Economics and Public Finance, Energy, Finance and Financial Sector, Foreign Trade and International Finance, Housing and Community Development, Labor and Employment, Social Welfare, Taxation, and Transportation and Public Works.

categories. These measures are available for each communication starting in 2023. We use all the tweets posted through the end of our sample period, yielding 158,745 tweets from 449 legislators.

3 Measurement

The empirical analysis requires a tweet-level measure of causal rhetoric. We define causal rhetoric as language that links the actual or potential action of a political actor to an outcome. We then distinguish whether this link is framed negatively, as blame, or positively, as credit. This definition leads to a two-step measurement problem: identifying whether a tweet contains a causal claim, and classifying the tone of that claim.

We implement this measure at scale by fine-tuning a RoBERTa-large model. We then validate the resulting classifications in three ways. First, we show that the classifier performs well against hand-coded labels. Second, we compare our measure with external benchmarks from America’s Political Pulse. Third, we document internal linguistic patterns consistent with the interpretation of blame and credit as distinct forms of causal rhetoric.

3.1 Definition

Our classification rests on two dimensions: causality and tone. A tweet is causal if it links the actual or potential action of a political actor to an outcome. Tone captures whether the tweet expresses a negative, neutral, or positive stance toward its subject.

Causality is binary. Political actors include politicians, political institutions, and politically aligned organizations. They exclude natural events, such as pandemics or disasters, and nonpolitical actors, such as scientific teams. A tweet is causal only if it links a politically meaningful action of such an actor to an outcome. The action may be explicit, as when the tweet refers to a particular bill, reform, or decision. It may also be implicit, as when the tweet attributes an outcome to a political actor while leaving the specific action unstated. In either case, the action may already have occurred, may be ongoing, or may be one the actor would take if in power. Outcomes may be economic, such as income or inflation, or social, such as civil liberties.

Tone captures the stance of the tweet toward its subject. We classify tone as negative, neutral, or positive. Negative and positive tone reflect unfavorable and favorable language, respectively. Neutral tone captures descriptive statements without evaluative content. Our manual annotation is designed to capture nuances that dictionary-based methods often miss.

Combining causality and tone yields our measure of rhetorical style. Tweets that are both causal and negative are coded as blame. Tweets that are both causal and positive are coded as credit. All remaining tweets, including noncausal tweets and causal tweets with neutral tone, are coded as none. Table A2 reports labeled examples.

3.2 Classification

To implement our measure of rhetorical style at scale, we begin by constructing a labeled dataset of 3,958 tweets, each annotated along the two dimensions introduced above: causality and tone. Because blame and credit tweets are likely to represent only a minority of the full corpus, a purely random sample would yield too few examples of the classes most relevant for training. This matters especially in our setting because fine-tuning performance depends on exposing the model to a sufficiently balanced set of examples across classes. To improve learning, we therefore combine random sampling with targeted oversampling, a common strategy in the classification literature (e.g., Davidson et al., 2017; He and Garcia, 2009; Talat and Hovy, 2016). Specifically, we first generate benchmark examples of blame and credit tweets for Democrats and Republicans using ChatGPT-4o, solely to guide sampling. We then compute cosine similarity between these benchmark examples and the full tweet corpus using SBERT-mini embeddings (Reimers and Gurevych, 2019). Half of the tweets selected for annotation are drawn from those most similar to the benchmarks, balancing by party and rhetorical style, while the remaining half are sampled at random.

To assess annotation reliability, we select a subsample of 530 tweets and have each of them independently labeled by all three coders. For this jointly annotated subsample, labels for causality and tone are assigned by majority vote, with ties broken randomly. The resulting protocol delivers strong inter-annotator agreement on the rhetorical style measure: the average pairwise correlation is 0.73 and Fleiss’ kappa is 0.64, typically interpreted as substantial agreement.⁷ Coders agree on rhetorical style in 67 percent of tweets, roughly six times the rate expected by chance.

We then fine-tune a RoBERTa-large model pre-trained on 154 million tweets (Loureiro et al., 2022).⁸ We split the labeled corpus into 80 percent training data and 20 percent validation data.⁹

⁷Fleiss’ kappa is a standard statistic for assessing agreement among more than two coders when observations are assigned to mutually exclusive categories. It is the natural multi-coder extension of Cohen’s kappa. Unlike simple percent agreement, it adjusts for the amount of agreement that would arise mechanically by chance, and therefore provides a more informative measure of annotation reliability.

⁸Fine-tuning adapts a pre-trained language model to a specific downstream task by updating its parameters on a smaller labeled dataset. In our setting, this allows the model to learn how causal attributions and tone are expressed in political language.

⁹We train for 10 epochs and retain the epoch with the highest F1 score on the validation set. As

The fine-tuned classifier assigns each tweet a probability distribution over the three classes—blame, credit, and none—and labels the tweet according to the highest probability class. Ambiguous cases, in which the model assigns similar probabilities to multiple classes, are extremely rare.¹⁰ 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, indicating strong performance by the standards of comparable text-classification tasks.

A natural concern is that our construction might miss a broader category of causal rhetoric that links political actions to outcomes without taking either a negative or a positive stance, and therefore extends beyond blame and credit. In practice, however, such neutral-toned causal tweets are exceedingly rare in our hand-labeled data, accounting for just 0.4 percent of observations. This suggests that, as a qualitative reading of political discourse would also imply, politicians almost never use causal rhetoric in a neutral way. Instead, when they make such causal claims, they typically do so to shape the attribution process through claims of blame or credit. To further validate this point, we train a separate classifier using only the binary causal label from the annotated dataset and find that its output is highly correlated with an indicator for whether a tweet is classified as either blame or credit ($\rho = 0.9$). Taken together, these patterns confirm that causal rhetoric in politics consists almost entirely of assigning blame and claiming credit.

3.3 Validation

We validate the measure against both external benchmarks and internal linguistic patterns. The external validation compares our labels with independent classifications from America’s Political Pulse. The internal validation asks whether tweets classified as blame and credit exhibit the semantic, syntactic, and emotional patterns one would expect from these categories.

External validation We begin with external validation relying on America’s Political Pulse, which classifies politicians’ communications into rhetorical categories, including policy-attack and credit-claiming, beginning in 2023 (Westwood and Lelkes, 2024). These categories are not identical to ours, but they provide a natural benchmark. Policy-attack is close to blame because both capture negative rhetoric directed at the actions of political actors. Credit-claiming is close to credit because both capture

robustness checks, we also fine-tune (i) a version of the model with half of its layers frozen and (ii) the well-known BERTweet model (Nguyen et al., 2020). Both alternatives perform slightly worse, but their outputs remain highly correlated with those of our baseline classifier, with Matthews correlation coefficients of 0.93 and 0.85, respectively.

¹⁰Figure A1 shows that, conditional on a predicted label, the associated probability distributions are strongly left-skewed and concentrated near 1, indicating that the classifier is rarely uncertain.

positive rhetoric that associates favorable outcomes or accomplishments with political actors.¹¹ For comparability, we restrict America’s Political Pulse to tweets, the same communication channel used in our analysis. This leaves 158,745 tweets through the end of our sample period.

The two measures align closely at the tweet level. For each tweet we can match across datasets, we compare our label with the corresponding America’s Political Pulse classification. Over the common sample period, this procedure yields 101,689 matched tweets. For blame, our label agrees with the policy-attack label in 87 percent of matched tweets, with a correlation of 0.62 between the two binary indicators. For credit, our label agrees with the credit-claiming label in 84 percent of matched tweets, with a corresponding correlation of 0.60.

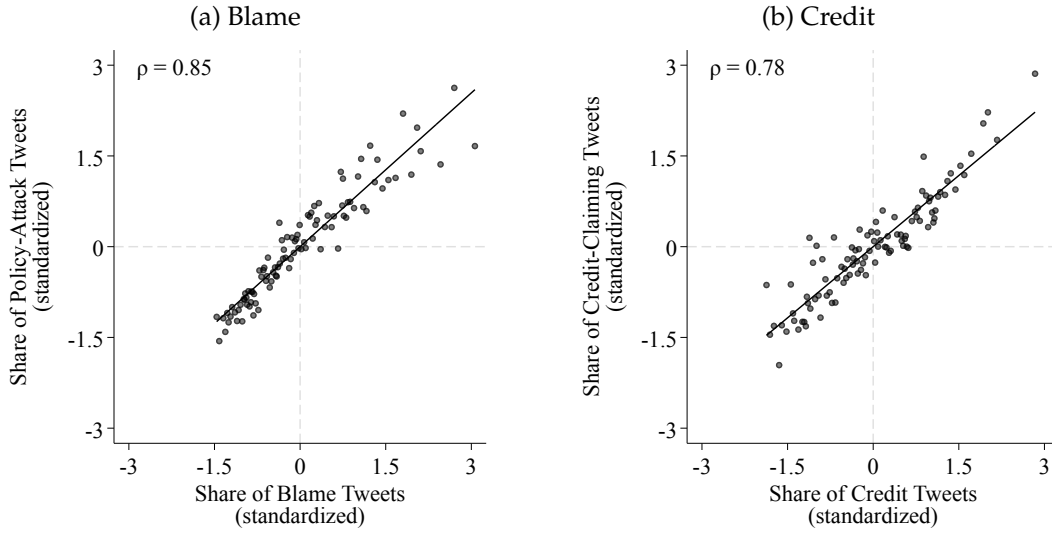
The measures also align in the cross-section of politicians. This validation is important because several of our empirical exercises use politician-level variation in the intensity of blame and credit. We restrict attention to the 443 politicians observed in both datasets over the common sample period, from the beginning of 2023 through the end of our sample. For each politician, we compute the share of tweets classified as policy-attack and credit-claiming in America’s Political Pulse and correlate these shares with our corresponding shares of blame and credit tweets. Figure 1 shows strong positive correlations: 0.85 between blame and policy-attack, and 0.78 between credit and credit-claiming. The tweet-level and politician-level comparisons therefore provide strong external support for the validity of our measure.

Internal validation We next turn to internal validation exercises, which help clarify the substantive content captured by our labels. We begin by examining the targets of causal claims. Syntactically, we measure whether tweets are self- or other-referential by comparing the relative frequency of second- and third-person pronouns to first-person pronouns. Figure A2a shows that blame tweets are more outward-directed, framing responsibility around others, whereas credit tweets are more self-directed. Semantically, we classify whether the attribution targets Democrats or Republicans.¹² Restricting attention to the 40 percent of tweets for which the target can be identified as either Democratic or Republican, Figure A2b shows that blame is disproportionately

¹¹In their codebook, [Westwood and Lelkes \(2024\)](#) define 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 on individuals involved with the policy, including accusations of lying or withholding information,” and credit-claiming as “communications about creating or passing legislation; securing government spending, grants, or funding; or emphasizing personal or party accomplishments in office.”

¹²Target classification is based on the Political DEBATE language model developed by [Burnham et al. \(2025\)](#). Political DEBATE is 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.

Figure 1: External Validation against America’s Political Pulse



Notes: This figure presents binned scatterplots comparing our politician-level measures of blame and credit with the corresponding politician-level measures from America’s Political Pulse (Westwood and Lelkes, 2024). In Panel (a), the horizontal axis reports the share of tweets classified as blame in our data, while the vertical axis reports the share of tweets classified as policy-attack in America’s Political Pulse. Panel (b) is analogous for credit and credit-claiming. Both sources are restricted to the common sample period, from the beginning of 2023 through the end of our sample. All variables are standardized. Observations correspond to politicians and are grouped into 100 bins. The correlation coefficient is reported in the upper-left corner of each panel.

directed at the opposing party, whereas credit is more often directed toward one’s own side.

We then examine the temporal orientation of causal rhetoric. Although our concept of causal rhetoric is not inherently temporal, blame and credit may differ in whether they connect actions to realized outcomes or to expected consequences. We capture this distinction using normalized counts of past and future tense. Figure A2c shows that blame is predominantly retrospective, whereas credit is more forward-looking, consistent with findings from psychology (Malle et al., 2014).

We next study the emotional content associated with each rhetorical style. Using a RoBERTa model fine-tuned for emotion detection in tweets (Camacho-Collados et al., 2022), we decompose the emotional profile of blame and credit tweets. Figure A3 shows that blame is dominated by anger, followed by disgust and fear, consistent with the idea that anger is directional and closely tied to causal attribution (Lazarus, 1991). Credit, by contrast, is mainly associated with optimism, reinforcing the earlier result that credit is more forward-looking.

Finally, we examine diagnostic language. Following Gentzkow and Shapiro (2010), we extract the bigrams that are most distinctive of each category. As shown in Figures A4 and A5, these bigrams closely match both intuition and the linguistic patterns documented above.¹³ We complement these diagnostics with illustrative tweets in

¹³For Democrats, “trump administration” is diagnostic of blame, while “will help” is diagnostic of

Tables A3 and A4.

As an additional check, our measure is not reducible to existing text-based measures of political language. Figure A6 compares blame and credit with sentiment, emotionality, and moral terminology, three dimensions commonly used to characterize political communication. The correlations are intuitive but limited: blame is more negative and credit more positive, but neither category is simply a relabeling of sentiment, emotional language, or moral language. This evidence supports the interpretation of blame and credit as a distinct dimension of political rhetoric.

This procedure yields the measure of causal rhetoric used throughout the rest of the paper. Starting from a definition of causal rhetoric as language linking political actions to outcomes, we fine-tune a language model to classify each tweet as blame, credit, or neither. The classifier performs well against hand-coded labels, aligns with an independent external benchmark, and displays the linguistic patterns expected of causal language in politics. We therefore use these measures to document the supply of causal rhetoric in congressional communication and to estimate its political returns.

4 The Supply of Causal Rhetoric

We use our measure of blame and credit to document three facts about the supply of causal rhetoric in congressional tweets. First, causal rhetoric is a large part of politicians’ communication, not a rare rhetorical device. Second, it is the language of political disagreement: when politicians assign blame or claim credit, the ruling party and the opposition diverge much more sharply in how they evaluate the political environment than they do in other tweets. Third, causal rhetoric has become substantially more common over time. These facts motivate the rest of the paper by showing that blame and credit are quantitatively important, tied to political disagreement, and increasingly prominent in elite political communication.

4.1 Prevalence

Causal rhetoric accounts for a substantial share of congressional communication on social media. In the full corpus, 16 percent of tweets are classified as blame and 20 percent as credit. Taken together, more than one third of congressional tweets assign blame or claim credit. This makes causal rhetoric a major margin of politicians’ communication rather than a narrow class of exceptional messages.

credit. For Republicans, “southern border” characterizes blame, while “will help” again identifies credit.

This prevalence is not evenly distributed across issues. Figure A7 shows that blame and credit are especially common in policy domains rather than sociocultural ones.¹⁴ Causal rhetoric is most prevalent in topics such as the economy, healthcare, immigration, and the environment, where political actors can most naturally be linked to social and economic outcomes, and less prevalent in topics such as gender, gun control, policing, and racial relations. This pattern is consistent with the interpretation of blame and credit as forms of causal attribution: politicians use them most where political actions can be connected to observable consequences.

Causal rhetoric is not concentrated among a narrow demographic type of politician. Figure A8 relates rhetorical style to politician characteristics. Restricting attention to demographic characteristics, age, gender, race, and education have little explanatory power for either blame or credit. Causal rhetoric therefore appears to be a broad feature of congressional communication rather than a style used only by a specific demographic group of politicians.

4.2 Disagreement

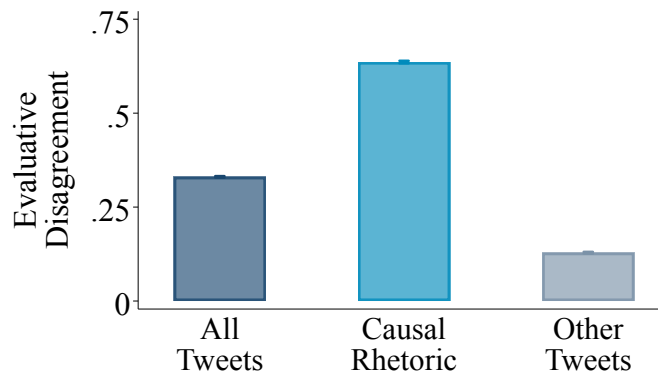
Causal rhetoric is also the language of political disagreement. We focus on a specific form of disagreement: divergence in how politicians evaluate the political environment. To measure this disagreement, we compare the average standardized sentiment of tweets posted by members of the ruling party with that of tweets posted by members of the opposition. This sentiment gap captures whether politicians on opposite sides of political control describe the state of the world in systematically different evaluative terms. We compute the gap separately for all tweets, tweets classified as blame or credit, and tweets classified as neither blame nor credit.

Figure 2 shows that the sentiment gap is concentrated in causal tweets. Across all tweets, the ruling-party-versus-opposition gap is 0.33 standard deviations. Among tweets classified as blame or credit, the gap nearly doubles, reaching 0.64 standard deviations. Among other tweets, it is only 0.13 standard deviations. Political disagreement is therefore not expressed equally across all communication. It is especially concentrated in causal rhetoric, where politicians connect political actions to outcomes through blame and credit.

This concentration of political disagreement in causal tweets is not mechanical. Blame and credit tweets, by construction, refer to political actors more often than non-causal tweets, so one concern is that they may exhibit larger sentiment gaps simply because they more often identify political targets. Figure A9 addresses this concern

¹⁴We classify tweets into eight broad topics using the Political DEBATE language model by Burnham et al. (2025): economy, environment, healthcare, and immigration on the policy side; and gender, gun control, policing, and racial relations on the sociocultural side.

Figure 2: Causal Rhetoric and Disagreement



Notes: The figure reports evaluative disagreement, measured as the difference in average standardized sentiment between tweets posted by members of the ruling party and tweets posted by members of the opposition. Higher values indicate a larger divergence in how politicians on opposite sides of political control evaluate the political environment. We compute this measure separately for all tweets, tweets classified as blame or credit, and tweets classified as neither blame nor credit. Bars represent 95 percent confidence intervals.

by restricting the analysis to tweets that can be clearly identified as targeting either Democrats or Republicans. The same pattern holds in this restricted sample, although the sentiment gaps are higher in levels. Causal rhetoric therefore appears to be the language of political disagreement, not merely a by-product of politicians mentioning political actors.

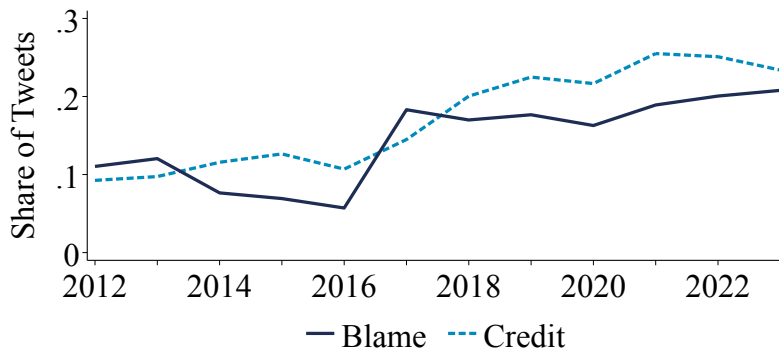
4.3 Trends

The prevalence of causal rhetoric, already large at the beginning of the sample, has grown over time. Figure 3 plots the yearly share of tweets classified as blame and credit from 2012 to 2023. Both dimensions rise substantially, from about 10 percent of tweets at the beginning of the sample to more than 20 percent by the end. Taken together, blame and credit grow from roughly one fifth to nearly one half of all congressional tweets.

The timing of the increase differs somewhat across the two rhetorical styles. Blame rises most sharply between 2016 and 2017, increasing by more than 10 percentage points in a single year. Credit follows a similar trajectory, although its growth is more gradual and concentrated between 2017 and 2020. By 2020, both series appear to plateau at high levels. Thus, an already large component of congressional communication becomes substantially more prominent over the sample period.

The rise does not appear to be driven by mechanical changes in the composition of the data. A first possibility is that different politicians are speaking, or that politicians are speaking about different issues. Figure A10 addresses this concern by estimating yearly changes in blame and credit while progressively adding politician and topic fixed effects, using the topic classification described above. The upward trend remains

Figure 3: Causal Rhetoric over Time



Notes: The figure presents the yearly share of tweets classified as blame and credit.

after accounting for both sets of fixed effects, indicating that the increase occurs within politicians and within topics rather than merely reflecting congressional turnover or shifts in the issues politicians discuss.

Additional checks address other mechanical explanations. The pattern is not driven by changes in sample coverage: Figure A11 shows similar trends when we exclude Senators, whose tweets enter our data only in 2017. It is also not a generic change in political language on Twitter: Figure A12 shows that sentiment, emotionality, and moral language do not exhibit comparable increases. Nor is the rise explained by Twitter platform changes. Appendix B shows no discontinuous increase around either the introduction of algorithmic feed curation in 2016 or the expansion of the character limit in 2017. Finally, Appendix C shows that the rise is not simply a Trump-style shift in congressional communication: Trump’s distinctive rhetorical feature is non-causal negative sentiment, not unusually high blame or credit.

Taken together, the supply evidence shows that causal rhetoric is a major and increasingly prominent component of politicians’ communication on social media. It is common, is the language of political disagreement, and increases over time. The next section therefore asks whether this rhetoric generates political returns.

5 The Political Returns of Causal Rhetoric

We next ask whether causal rhetoric generates political returns. We interpret politicians’ use of blame and credit on social media as a rhetorical production problem (Ariador et al., 2024): politicians choose rhetorical inputs partly based on the political returns those inputs can generate. This section studies those returns by asking whether exposure to politicians assigning blame and claiming credit increases political support.

We study this question using two complementary sources of evidence. The first

is small campaign contributions. Although not perfectly representative of the general population, small donations can still provide a revealed-preference measure of political support that can be linked systematically to politicians' communication in our data (Bouton et al., 2022; Petrova et al., 2021). They also allow us to distinguish two margins of support: whether rhetoric mobilizes additional donors, and whether it deepens support by increasing average giving among donors.

Using a cross-county shift-share-style exposure design that combines politician-month rhetoric with quasi-exogenous geographic variation in Twitter adoption (Müller and Schwarz, 2023), we estimate the effect of exposure to politicians' blame- and credit-intensive communication on small-donor support. The design asks whether counties more exposed to Twitter respond more to the same politician in months when that politician uses more blame or credit, rather than considering a timing comparison around individual tweets. The estimates show asymmetric returns. Blame has the larger aggregate return, mainly by expanding the donor base. Credit generates much smaller aggregate returns, with detectable effects only on average giving. These effects also differ across donor types: ideologically extreme donors respond more to blame, while ideologically moderate donors respond more to credit.

The second source of evidence is survey data on voting intentions in the 2020 House elections. The survey analysis complements the donation design in three ways. It measures support directly as intended vote choice, covers respondents beyond the selected population of donors, and uses a short-run timing comparison.

Using an event-study design around blame-only and credit-only days before the election, we examine whether support for a politician changes after that politician's Twitter communication consists entirely of one rhetorical style. Support increases in the week after blame-only days, especially among respondents who report using social media for political news. Credit-only days show no comparable effect. The survey evidence therefore points in the same direction as the donation evidence while also supporting the social-media exposure channel.

We close by examining amplification. Retweet data show that blame travels farther than other rhetorical styles, identifying reach as a channel consistent with blame's larger political returns.

5.1 Empirical Strategy for Donations

We estimate the effect of causal rhetoric on small-donor support using a shift-share exposure design that relies on quasi-exogenous exposure shares (Borusyak et al., 2025; Goldsmith-Pinkham et al., 2020). The shifts are politician-month rhetoric: in a given month, a politician uses more or less blame or credit. The exposure share is county-

level exposure to Twitter: residents of counties with higher Twitter adoption are more likely to encounter the same politician’s Twitter communication. Identification comes from within-politician-month comparisons: we ask whether counties with greater quasi-exogenous Twitter exposure donate more to the same politician in months when that politician uses more blame or credit, net of persistent politician-county ties and county-month shocks.

We use the geographic detail in campaign finance records to construct a politician-by-county-by-month panel. For each politician, county, and month in which the politician tweets, we measure three outcomes from small donors: total donation revenue, the number of donors, and the average amount contributed per donor. We estimate:

$$y_{icm} = \beta_1(\text{Blame}_{im} \times \text{Users}_c) + \beta_2(\text{Credit}_{im} \times \text{Users}_c) + \gamma X_{icm} + \lambda_{ic} + \mu_{im} + \eta_{mc} + \varepsilon_{icm}. \quad (1)$$

Here, y_{icm} is the outcome for politician i in county c and month m : $\log(1+\text{donation revenue})$, $\log(1+\text{number of donors})$, or $\log(1+\text{average per donor})$. Blame_{im} and Credit_{im} denote the standardized shares of blame and credit tweets posted by politician i in month m . Users_c measures county-level Twitter use, defined as the log of one plus the number of Twitter users in county c . The vector X_{icm} denotes controls. In our preferred specification, it includes $\text{Volume}_{im} \times \text{Users}_c$, where Volume_{im} is the standardized total number of tweets posted by politician i in month m . This separates exposure to rhetorical style from exposure to overall political activity on Twitter.

The fixed effects make the exposure comparison demanding. Politician-by-month fixed effects (μ_{im}) absorb the politician-month rhetorical shift itself, as well as any month-specific shock to politician i , including changes in popularity, fundraising effort, or national visibility that affect donations across all counties. Politician-by-county fixed effects (λ_{ic}) absorb persistent differences in politician i ’s support across counties, including geographic ties between politicians and specific places. County-by-month fixed effects (η_{mc}) absorb local shocks to donations in month m that are common across politicians, such as county-level changes in political engagement or economic conditions. The coefficients are therefore identified by comparing, within the same politician-month, whether donations respond more in counties with higher Twitter exposure, net of persistent politician-county links and county-month shocks.

The main identification challenge is that county Twitter adoption may be endogenous to local political and economic conditions. Counties with greater partisan competition, higher donor capacity, or different unobservable characteristics may exhibit different levels of Twitter adoption and may also differ systematically in political engagement. If so, the interaction between politicians’ rhetoric and county Twitter use could

partly capture differences in the local demand for political communication rather than the causal effect of exposure to that rhetoric. To address this concern, we exploit quasi-exogenous variation in Twitter adoption induced by the platform’s early diffusion after the South by Southwest (SXSW) festival in 2007, a well-documented shock that accelerated Twitter’s initial spread in the United States (Fujiwara et al., 2024; Müller and Schwarz, 2023). Specifically, we instrument county-level Twitter adoption with the number of followers of the official SXSW Twitter account who joined Twitter in 2007. The instrument isolates variation in local Twitter use generated by early exposure to the platform, rather than by contemporaneous local determinants of political donations.¹⁵

In practice, we instrument $Users_c$ with $\log(1 + SXSWFollowers2007_c)$, where $SXSWFollowers2007_c$ denotes the number of followers of the official SXSW account in county c who joined Twitter in 2007, after the SXSW Festival. We estimate a standard two-stage least squares specification, instrumenting each interaction in Equation 1 that involves county Twitter adoption with the corresponding interaction based on $SXSWFollowers2007_c$. Following Fujiwara et al. (2024); Müller and Schwarz (2023), we also interact our main regressors of interest with the number of SXSW followers who joined Twitter before 2007, at any point in 2006. This ensures that identification comes from comparing counties that are similar in their pre-existing propensity to engage with SXSW, and therefore more comparable in their baseline relationship to the platform.

The identifying assumption is a share-exogeneity condition. Conditional on the fixed effects, tweet-volume exposure, and pre-2007 SXSW propensity controls, counties with greater SXSW-induced Twitter adoption would have experienced similar changes in politician-specific donations across more and less blame-intensive, or more and less credit-intensive, months absent differential exposure to politicians’ Twitter communication. The design does not require politicians’ rhetorical choices to be randomly timed: politicians may choose when to assign blame or claim credit in response to political events. Identification comes instead from the SXSW-induced component of county Twitter exposure, which must be unrelated to unobserved county-level donation responses to politician-month rhetorical shifts except through exposure to the politician’s tweets. Under this assumption, the estimates identify the causal effect of exposure to more blame- or credit-intensive politician-month communication periods on small-donor support.

¹⁵This instrument is now widely used in the literature (e.g., Boken et al., 2023). We refer to the original papers for detailed evidence on both its relevance for the diffusion of local Twitter activity and its plausible exogeneity.

Table 1: Donation Returns

	(1)	(2)	(3)
	Revenue from Donations	Number of Donors	Average per Donor
Blame x Users	0.005*** (0.001)	0.003*** (0.000)	0.003*** (0.001)
Credit x Users	0.001*** (0.000)	0.000** (0.000)	0.001*** (0.000)
Blame x SXSWFollower2006	-0.006 (0.004)	-0.002 (0.002)	-0.004 (0.003)
Credit x SXSWFollower2006	0.003 (0.002)	0.001 (0.001)	0.002* (0.001)
Politician x Month FE	✓	✓	✓
Politician x County FE	✓	✓	✓
County x Month FE	✓	✓	✓
Tweet Volume Control	✓	✓	✓
Observations	168,232,932	168,232,932	168,232,932
Clusters	3,108	3,108	3,108
F statistic	63.65	63.65	63.65

Notes: The table presents 2SLS estimates of Equation 1. The outcomes are log-donation revenue in column (1), log-number of donors in column (2), and log-average donation per donor in column (3). The main regressors are the standardized monthly shares of blame and credit tweets posted by the politician, each interacted with the log of one plus the number of Twitter users in the county. The table also reports the coefficients on the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2006. Interactions with county Twitter users are instrumented using the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2007. All specifications include politician-by-month, politician-by-county, and county-by-month fixed effects, as well as the interaction between tweet volume and county Twitter users. Standard errors in parentheses are clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

5.2 Results for Donations

Exposure to causal rhetoric increases small-donor support, but the returns are asymmetric. Table 1 reports estimates of Equation 1 using log total small-donation revenue as the outcome in column (1).¹⁶ A one-standard-deviation increase in the monthly share of blame tweets increases donation revenue by roughly 2.6 percent in a county with average Twitter adoption. The corresponding increase for credit is about 0.5 percent. Blame therefore generates the larger aggregate return.

These aggregate returns operate through different margins of support. Donation revenue can increase because rhetoric mobilizes additional donors, raising the number of donors, or because it deepens support among donors, raising the average amount contributed per donor. To distinguish between these margins, Table 1 use the log number of donors and log average donation per donor as outcomes in columns (2) and (3). Blame is the only rhetorical style that expands the donor base. A one-standard-deviation increase in blame increases the number of donors by about 1.6 percent in

¹⁶Table A6, Table A7, and Table A8 report the corresponding first-stage, OLS, and reduced-form results.

a county with average Twitter adoption, while credit has essentially no effect on this margin. Credit instead has a smaller effect on support deepening: a one-standard-deviation increase in credit raises the average donation per donor by about 0.5 percent in a county with average Twitter adoption. The corresponding estimates for blame on average donation size are positive but less precise across some robustness checks. Overall, blame is effective at mobilizing additional donors, while credit only slightly increases average giving among those already willing to donate.

Persuasion rate To benchmark the magnitude of the donor-mobilization effect, we compute persuasion rates following [DellaVigna and Gentzkow \(2010\)](#), focusing on the donor-count outcome.¹⁷ The implied persuasion rate is 0.94%.¹⁸ This magnitude is comparable to estimates for political advertising ([Spenkuch and Toniatti, 2018](#)) and opening a Twitter account ([Petrova et al., 2021](#)), though smaller than the effects associated with virality ([Boken et al., 2023](#)) and the average persuasion rates summarized by [DellaVigna and Gentzkow \(2010\)](#).

Heterogeneity by donor ideology The donation returns to blame and credit differ not only by margin of support, but also by donor ideology. The aggregate results show that blame is more effective at mobilizing additional donors, while credit slightly increases average giving. We next ask whether these two rhetorical styles activate different ideological constituencies within the donor base.

To study this heterogeneity, we use the donor-level ideology measure constructed by [Bonica \(2014\)](#) from campaign finance records, which is analogous to NOMINATE scores for legislators. Within each electoral cycle, we classify donors as moderate or extreme depending on whether the absolute value of their ideological score falls below or above the cycle-specific median. We then re-estimate Equation 1 separately by donor type.

Table 2 reveals a clear ideological divide. Moderate donors respond positively to credit and negatively to blame. Extreme donors display the opposite pattern: they respond positively to blame and negatively to credit. This symmetry is qualitative rather than quantitative. The positive response of extreme donors to blame is substantially larger than the positive response of moderate donors to credit. These differences indicate that blame and credit do not merely operate through different margins of support; they also activate different ideological constituencies within the donor base. The two

¹⁷As emphasized by [Boken et al. \(2023\)](#), persuasion rates are most naturally interpreted when the underlying outcome is close to a binary participation decision.

¹⁸Appendix D details the construction of this estimate and shows that it should be interpreted as a lower bound.

Table 2: Donation Returns by Donor Ideology

	(1)	(2)	(3)	(4)	(5)	(6)
	Moderate Donors			Extreme Donors		
	Revenue from Donations	Number of Donors	Average per Donor	Revenue from Donations	Number of Donors	Average per Donor
Blame x Users	-0.003*** (0.001)	-0.002*** (0.000)	-0.001*** (0.000)	0.010*** (0.002)	0.005*** (0.001)	0.006*** (0.001)
Credit x Users	0.005*** (0.001)	0.002*** (0.000)	0.003*** (0.000)	-0.003*** (0.001)	-0.002*** (0.000)	-0.002*** (0.000)
Blame x SXSWFollower2006	-0.001 (0.003)	-0.000 (0.001)	-0.000 (0.002)	-0.005 (0.006)	-0.002 (0.003)	-0.003 (0.004)
Credit x SXSWFollower2006	0.000 (0.003)	-0.000 (0.001)	0.000 (0.002)	0.003 (0.002)	0.001 (0.001)	0.002* (0.001)
Politician x Month FE	✓	✓	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓	✓	✓
Tweet Volume Control	✓	✓	✓	✓	✓	✓
Observations	168,232,932	168,232,932	168,232,932	168,232,932	168,232,932	168,232,932
Clusters	3,108	3,108	3,108	3,108	3,108	3,108
F statistic	63.65	63.65	63.65	63.65	63.65	63.65

Notes: The table presents 2SLS estimates of Equation 1 separately by donor ideology. Columns (1)–(3) use outcomes computed from ideologically moderate donors, and columns (4)–(6) use outcomes computed from ideologically extreme donors. Within each donor group, the outcomes are log-donation revenue, log-number of donors, and log-average donation per donor. The main regressors are the standardized monthly shares of blame and credit tweets posted by the politician, each interacted with the log of one plus the number of Twitter users in the county. The table also reports the coefficients on the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2006. Interactions with county Twitter users are instrumented using the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2007. All specifications include politician-by-month, politician-by-county, and county-by-month fixed effects, as well as the interaction between tweet volume and county Twitter users. Standard errors in parentheses are clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

rhetorical styles are therefore politically complementary despite the larger aggregate returns to blame.

Robustness checks We estimate several additional specifications that address alternative interpretations of the donation results. First, the estimated returns to blame and credit may reflect exposure to negative and positive sentiment rather than exposure to causal attribution. This would be a concern if politicians systematically use blame when they also use more non-causal negative rhetoric, or credit when they also use more non-causal positive rhetoric. In that case, the estimates could partly capture the returns to affective sentiment rather than the returns to causal rhetoric. To address this concern, we control for the shares of non-causal tweets with negative sentiment and non-causal tweets with positive sentiment, each interacted with county Twitter adoption and instrumented with the corresponding SXSW interaction. The results, reported in Table A9, are essentially unchanged.

Second, the estimates may reflect exposure to partisan attack and praise rather than exposure to causal rhetoric. This would be a concern if politicians use blame and credit disproportionately during moments of partisan conflict, when they are also

more likely to criticize the opposing party or praise their own side. In that case, the estimates could partly capture the returns to partisan attack or praise rather than the returns to assigning blame or claiming credit. We therefore construct measures of non-causal attacks and non-causal praise. Non-causal attacks are tweets that are non-causal, negative in tone, and about the opposing party; non-causal praises are tweets that are non-causal, positive in tone, and about the politician’s own party. We then control for the shares of these tweets, interacted with county Twitter adoption and instrumented with the corresponding SXSW interactions.¹⁹ The results, also reported in Table A9, are essentially unchanged.

Third, blame and credit may coincide with other politically salient content. Politicians may assign more blame or claim more credit when economic news is released, scandals emerge, or other salient information becomes public. If county responsiveness to this accompanying content is correlated with SXSW-induced Twitter adoption through demographics, media-consumption patterns, or socioeconomic conditions, the estimates could partly capture differential responses to correlated political events rather than the rhetorical margin itself. To address this concern, we estimate a richer specification that allows the effects of blame, credit, and tweet volume to vary flexibly with county characteristics. We use the 29 county-level covariates employed by Fujiwara et al. (2024), summarizing them into eight principal components to avoid a high-dimensional set of interactions, and interact these principal components with blame, credit, and tweet volume.²⁰ This specification absorbs differential responsiveness associated with observable county characteristics, so correlated political events would have to be differentially amplified through the SXSW-induced Twitter-exposure margin over and above these demographic, socioeconomic, media-consumption, and political-engagement channels. The results, also reported in Table A9, are essentially unchanged, except for blame becoming noisier on average donation.

Finally, the log transformations may be sensitive to the large number of zeros in the donation outcomes (Chen and Roth, 2024). We therefore re-estimate the same specification, as well as the previous robustness checks, using a binary outcome equal to one if a politician receives at least one donation from a given county in a given month, and zero otherwise. The results, reported in Table A10, are consistent with the donor-count evidence: the effects point in the same direction even when we use a coarser

¹⁹To identify whether tweets are about the opposing party or the politician’s own party, we use the same target-classification procedure employed in the validation exercises, based on the Political DEBATE language model by Burnham et al. (2025).

²⁰The 29 covariates include demographic and socioeconomic characteristics, as well as measures of media exposure such as prime-time television viewing and Fox News viewership. Table A5 lists the underlying covariates. We retain principal components using the Kaiser criterion, keeping components with eigenvalues greater than one. This yields eight components, which together explain 74% of the variance in the 29 county-level covariates.

and more restrictive measure of donor mobilization that captures only whether the politician receives any donation from a county-month, rather than how many donors contribute.

Overall, the donation evidence points to asymmetric political returns. Blame increases small-donor support primarily by mobilizing additional donors. Credit generates smaller returns and only increases average giving among donors. These results establish the donation margin of political support. We next ask whether the same pattern appears in a broader survey measure of intended vote choice.

5.3 Empirical Strategy for Voting Intentions

We estimate whether causal rhetoric changes voting intentions using an event-study design around days when a politician’s Twitter communication consists entirely of a single rhetorical style. For each politician, we define a blame-only day as a day on which all the politician’s tweets are classified as blame; credit-only days are defined analogously. This definition isolates days with a clear rhetorical style: the politician uses only blame or only credit. The identifying comparison asks whether support for the politician changes after blame-only or credit-only days, net of district-month conditions, date shocks, partisan alignment, and respondent characteristics.

We use Nationscape survey responses on voting intentions in the 2020 House elections. Our main outcome, Support_{idt} , is an indicator equal to one if respondent i in district d , interviewed on date t , reports intending to vote for the politician, and zero otherwise. We link each respondent to the Twitter communication of the elected representative serving the respondent’s congressional district. We then compare respondents interviewed in the seven days after blame-only or credit-only days with respondents interviewed during the six days before the rhetoric-only day and on the rhetoric-only day itself. Respondents interviewed on the rhetoric-only day itself are included in the comparison group because we do not observe whether the survey response occurred before or after the politician posted on Twitter.

Formally, for each rhetorical style $r \in \{\text{Blame}, \text{Credit}\}$, we construct an indicator $\text{Post-}r_{dt}$ equal to one if respondent i is interviewed in district d during the seven days after a rhetoric-only day of style r by the district’s politician. The comparison group consists of respondents interviewed during the six days before the rhetoric-only day and on the rhetoric-only day itself. We exclude observations that fall in both the pre- and post-period for the same rhetorical style. We then estimate:

$$\text{Support}_{idt} = \beta \text{Post-}r_{dt} + \lambda_{d \times m(t)} + \mu_t + \eta_{a(i,d)} + \theta_{X_i} + \varepsilon_{idt}, \quad (2)$$

where Support_{idt} is the voting-intention outcome; $\lambda_{d \times m(t)}$ are district-by-month fixed effects; μ_t are date fixed effects; $\eta_{a(i,d)}$ are fixed effects for respondent–politician partisan alignment, defined by respondent party identification interacted with an indicator for whether the politician is Republican; and θ_{X_i} denotes fixed effects for respondent gender, age group, race, education, and employment status.²¹ Standard errors are clustered at the congressional-district level.

The identifying assumption is that, within a district and month, the exact timing of survey interviews is as-good-as-random with respect to the politician’s blame-only or credit-only days, after accounting for date effects, partisan alignment, and respondent demographics. District-by-month fixed effects compare respondents in the same district and month, holding fixed local political conditions and the local campaign environment. Date fixed effects absorb national shocks common to all districts. Party-identification-by-politician-party fixed effects capture partisan alignment by allowing baseline support to differ across respondent party groups depending on whether the politician is Democratic or Republican. Respondent-characteristic fixed effects absorb differences in support across observable demographic groups.

The balance checks support this timing comparison. Table A11 shows that observable demographic characteristics of respondents interviewed before rhetoric-only days are statistically indistinguishable from those of respondents interviewed after rhetoric-only days. Table A12 shows the same for partisan alignment, measured by the interaction between respondent party identification and the politician’s party. These checks reduce the concern that the before-after comparison is driven by changes in the composition of respondents within district-months.

5.4 Results for Voting Intentions

Blame increases voting support in the short run, while credit produces a smaller and imprecise effect. Table 3 reports estimates of Equation 2 for all respondents in column (1). Support for the politician rises by 1 percentage point in the week after blame-only days, and the estimate is precise. The corresponding estimate after credit-only days is smaller, at 0.6 percentage points, and is not statistically significant. Blame therefore generates the clearer voting-intention response.

The dynamic pattern supports this timing interpretation. In Figure A13, we estimate a dynamic version of Equation 2 using three-day bins around rhetoric-only days to ensure sufficient precision. Support is flat in the two weeks before blame-only days: the pre-period coefficients are small and insignificant. After blame-only days, support

²¹Respondent party identification is coded in four categories: Democrat, Republican, Independent, and something else.

Table 3: Voting Intentions

	(1)	(2)	(3)
	All Respondents	Split by Social Media Use	
		No Social Media Use	Social Media Use
(a) Blame			
Post-Blame	0.010** (0.005)	0.006 (0.010)	0.014** (0.006)
District x Month FE	✓	✓	✓
Date FE	✓	✓	✓
Partisan Alignment FE	✓	✓	✓
Respondent Covariate FE	✓	✓	✓
Observations	20,154	5,781	14,300
Clusters	281	275	281
(b) Credit			
Post-Credit	0.006 (0.004)	-0.000 (0.009)	0.006 (0.005)
District x Month FE	✓	✓	✓
Date FE	✓	✓	✓
Partisan Alignment FE	✓	✓	✓
Respondent Covariate FE	✓	✓	✓
Observations	29,109	8,257	20,756
Clusters	325	322	325

Notes: The table presents estimates of Equation 2. The outcome is an indicator equal to one if the respondent reports intending to vote for the politician, and zero otherwise. Panel (a) reports estimates for blame-only days, and Panel (b) reports estimates for credit-only days. The treatment indicator equals one for respondents interviewed in the seven days after a blame-only or credit-only day. Column (1) uses all respondents. Column (2) restricts the sample to respondents who do not report having seen or heard political news on social media in the past week. Column (3) restricts the sample to respondents who do report having seen or heard political news on social media in the past week. All specifications include district-by-month, date, partisan alignment, and respondent covariate fixed effects. Standard errors in parentheses are clustered at the congressional-district level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

increases in the first post-period bins and fades within two weeks. The corresponding estimates for credit-only days are smaller and show no systematic post-period increase. The support response therefore appears after blame-only days rather than reflecting pre-existing movement in support.

Social media channel The survey evidence also supports the social-media exposure channel. If the response operates through Twitter exposure to politicians' communication, it should be stronger among respondents more likely to encounter political news on social media. To test this prediction, we re-estimate Equation 2 separately by whether respondents report having seen or heard news about politics on social media in the past week. Table 3 show that the response is concentrated among these respondents in columns (2) and (3). Among social-media news users, support for the politician increases by 1.5 percentage points after blame-only days. Among respon-

dents who do not report using social media for political news, the estimated effect is 0.4 percentage points and is not precisely estimated. The corresponding estimates for credit-only days are positive but imprecise. This heterogeneity reinforces the interpretation that the voting-intention response operates through exposure to politicians' communication on social media.

Robustness checks We estimate several additional specifications to assess the robustness of the survey results. These checks address alternative interpretations related to time-varying political shocks, Twitter activity around rhetoric-only days, and the definition of rhetoric-only days.

A first concern is that national or state-level events may shift both politicians' rhetoric and voter support differently across politicians' parties. Date fixed effects absorb shocks common to all districts, but they do not absorb shocks that differentially affect districts represented by Democratic and Republican politicians on the same day. For example, an event could lead Democratic politicians to use more blame while also increasing support for Democratic politicians for reasons unrelated to their own Twitter communication. To address this concern, we replace date fixed effects with date-by-politician-party fixed effects, which absorb daily shocks that differ by the politician's party. The results, reported in Table A13, are essentially unchanged.

A second related concern is that events may shift support differently across respondent party groups while also changing the timing of blame-only or credit-only days. For example, an event could increase support among Democratic respondents and coincide with blame-only days by politicians in districts where Democratic respondents are more common. To address this concern, we replace date fixed effects with date-by-respondent-party-identification fixed effects, which absorb daily shocks that differ across respondent party groups. The results, also reported in Table A13, are essentially unchanged.

A third concern is that rhetoric-only days may differ from nearby comparison days not only in rhetorical content, but also in the politician's Twitter activity on that day. If the timing of blame-only or credit-only days is correlated with unusual daily tweet volume, the estimated response could partly reflect that Twitter activity rather than the rhetorical style itself. To address this concern, we re-estimate Equation 2 absorbing fixed effects for the number of tweets posted on the rhetoric-only day.²² The results, reported in Table A14, are essentially unchanged.

A related concern is that rhetoric-only days may occur during periods of unusual communication activity more broadly, even if the rhetoric-only day itself is not un-

²²When multiple rhetoric-only days occur in the comparison window, we assign the maximum raw daily tweet volume among those days.

usual in volume. In that case, the estimates could reflect broader campaign intensity or political salience in the surrounding window, rather than voters' response to blame or credit rhetoric. To address this concern, we control for the log of total tweet volume across the full comparison window. The results, also reported in Table A14, are essentially unchanged.

We also test whether the results depend on our stringent definition of rhetoric-only days. The baseline specification defines a blame-only day as one on which all tweets posted by the politician are classified as blame, with credit-only days defined analogously. We re-estimate the survey specification using an alternative threshold based on the politician-specific 99th percentile of daily blame and credit intensity. This definition captures days on which a politician's communication is more concentrated in blame or credit than usual. The alternative definition is broader, but also potentially noisier, because daily rhetoric intensity is constructed from tweet-level binary indicators and can therefore be sensitive to small daily tweet counts. The results, reported in Table A15, are qualitatively unchanged across specifications, though somewhat attenuated. This attenuation is consistent with the broader and noisier definition, which includes less rhetorically clean communication and therefore weakens the contrast between treated and comparison periods.

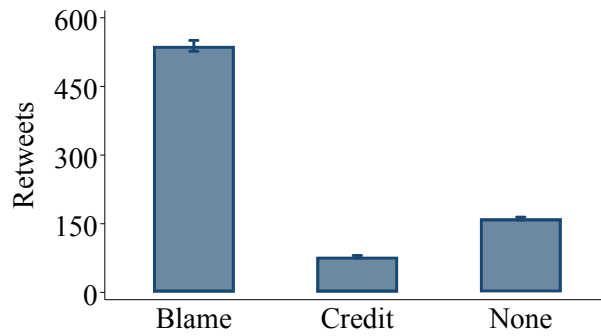
Taken together, the survey results extend the donation evidence beyond selected donors, study short-run timing directly, and support the social-media exposure channel. Political support increases in the week after blame-only days in a broader survey population, especially among respondents more likely to encounter political communication on social media. Credit shows no comparable pattern. These results reinforce the interpretation of blame as the rhetorical style most effective at generating political support.

5.5 Blame and Amplification

The preceding results show that blame is the rhetorical style most consistently associated with political support: it raises donation revenue by expanding the donor base and increases voting support among respondents more likely to encounter political communication on social media. We close by examining a possible amplification mechanism. If blame travels farther on Twitter, it may generate larger returns partly because more users encounter it.

We therefore examine whether blame is disproportionately associated with retweet activity. Retweets measure diffusion rather than endorsement. They may reflect agreement, criticism, or attention, but they are directly informative about reach, the relevant margin for assessing amplification. Retweet data are available for roughly half of the

Figure 4: Blame and Retweet Amplification



Notes: The figure reports the average number of retweets received by tweets classified as blame, credit, and neither blame nor credit, restricting to tweets with non-missing retweet data. Bars represent category means. Bars represent 95 percent confidence intervals.

tweets in our corpus and only through March 2020. Table E1 shows that this engagement subsample is broadly representative of the underlying tweet sample, reducing concerns that the results are driven by selection into non-missing retweet data. Appendix E provides details on the data and empirical specifications.

Blame receives disproportionate attention. In the sample with non-missing retweet data, blame tweets account for only 15% of tweets but nearly 40% of all retweets. Figure 4 shows they receive, on average, nearly four times as many retweets as tweets in the other rhetorical categories. Thus, conditional on being posted, blame is much more likely to circulate widely.

More structured analyses show the same pattern. Figure E1 shows that, relative to tweets not classified as blame or credit, blame tweets receive about 0.15 standard deviations more retweets after controlling for politician-by-date and topic-by-date fixed effects. Credit tweets receive no comparable retweet premium. Figure E2 shows that this relationship is strongest in the upper tail of the retweet distribution: blame is negatively associated with appearing in the lower deciles of retweet counts and increasingly positively associated with appearing in the upper deciles, with the largest estimate in the top decile.

These findings point to amplification as a channel through which blame can generate larger political returns. Retweets measure diffusion rather than endorsement, and on that margin blame travels farther than other rhetorical styles. This greater reach offers a plausible explanation for why blame is especially effective at generating political support, and connects our evidence to Boken et al. (2023), who show that small donations rise sharply when politicians go viral.

6 Interpreting Economic Causal Rhetoric

The preceding sections show that causal rhetoric is widely used in congressional communication and that its political returns are concentrated in blame. This raises a natural interpretive question: what kind of claims does this rhetoric contain? Answering this question can help clarify whether causal rhetoric helps voters assess the mapping between political actions and outcomes, or instead supplies politically convenient claims about that mapping. We focus on the economic domain, where causal rhetoric is common and where claims about political actions and outcomes can be assessed against standard macroeconomic reasoning. The evidence points to underqualified yet plausible, politically convenient claims.

This section develops this interpretation using two descriptive exercises. First, we assess economic causal claims at the level of support-as-stated. The content analysis shows that these claims are usually underqualified: most require qualifications that the tweet does not provide, especially when the claim assigns blame. At the same time, it suggests that these claims may remain plausible rather than outlandish: outright unsupported claims are rare. Second, we study whether economic causal rhetoric moves with economic bill introductions. The bill analysis reinforces the plausibility of these claims, because rhetoric lines up with observable political action, and it also shows their political convenience: blame rises around opposing-party economic bills, while credit rises around own-party economic bills. Together, the two exercises suggest that causal rhetoric, and especially blame, can turn complex economic mechanisms into underqualified yet plausible, politically convenient claims.

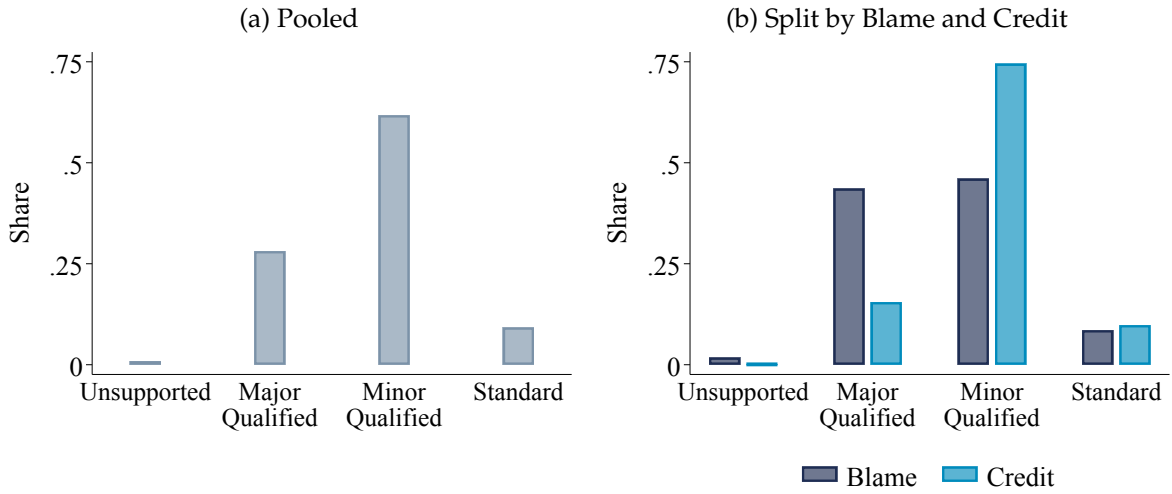
6.1 Economic Claims as Stated

Assessing the grounding of causal claims is difficult because the object of interest is not a simple factual statement. A claim about inflation, employment, or government spending may be supported under some assumptions, misleading under others, or valid only once timing, incidence, magnitude, and general-equilibrium effects are specified. Many claims therefore do not fit naturally into a true-false classification. We instead classify each claim according to the extent to which it is supported by standard macroeconomic reasoning as stated.

We analyze 413,836 tweets that are both classified as blame or credit and identified by our topic classifier as being about the economy. For each tweet, we use an AI-assisted coding procedure to extract the dominant causal claim, identify the implied economic mechanism, and classify the claim’s level of support.²³ The classification

²³We implement this procedure using GPT-5 mini under a fixed codebook and prompt that instruct

Figure 5: Support Level for Economic Causal Rhetoric



Notes: Panel (a) reports the distribution of macroeconomic-support labels across tweets classified as blame or credit and identified by our topic classification as being about the economy. Panel (b) reports the same distribution separately for tweets classified as blame and credit.

assigns each claim to one of four categories.

A claim is *supported* if standard macroeconomic reasoning supports it under ordinary assumptions, without requiring important additional qualifications. A claim is *minor qualified* if the basic mechanism and direction are standard, but the claim requires familiar caveats about context, timing, magnitude, or incidence. A claim is *major qualified* if some route to support exists, but only after introducing substantial unstated conditions or offsetting forces. A claim is *unsupported* if no ordinary mainstream macroeconomic route supports the claim under its baseline interpretation. The resulting measure therefore captures how much qualification is needed before the causal claim can be reconciled with standard macroeconomic reasoning.

Most economic causal claims require qualification. Figure 5a reports the distribution of support labels across economic blame and credit tweets. Only about 10 percent of claims are classified as supported as stated. Roughly 90 percent require either minor or major qualifications. The largest category is minor qualified, at 62 percent, followed by major qualified, at 28 percent. Unsupported claims are rare, accounting for less than 1 percent of observations.

The need for qualification is especially pronounced for blame. Figure 5b reports the same classification separately for blame and credit tweets. Blame claims are much more likely to be major qualified: 44 percent of blame claims fall in this category, compared with 16 percent of credit claims. Credit claims are instead more concentrated

the model to act as a careful research coder and evaluate claims using standard mainstream macroeconomic reasoning. Appendix F provides additional details on the prompt, codebook, and implementation.

in the minor qualified category, which accounts for 75 percent of credit claims and 46 percent of blame claims. Unsupported claims remain rare for both styles, but they are more common among blame tweets: 1.8 percent of blame claims are unsupported, compared with 0.17 percent of credit claims.

The content analysis delivers the first part of the interpretation. Economic causal claims are usually underqualified: politicians rarely state them in a form that standard macroeconomic reasoning supports without qualification, especially when the claim assigns blame. At the same time, these claims are not usually outlandish: politicians almost never make claims that standard reasoning rejects outright. Most claims instead point to mechanisms that can be made economically plausible, but only after adding conditions that the rhetoric itself does not state. The next subsection asks whether this plausibility also appears in the use of rhetoric around observable political actions, and whether the resulting claims are politically aligned in a convenient direction.

6.2 Economic Claims around Economic Bills

The content analysis shows that economic causal rhetoric often occupies a middle ground: claims are underqualified, but rarely unsupported outright. We now ask whether the use of this rhetoric is consistent with that middle ground. Congressional bill introductions provide a concrete political action taken by a specific party. If economic causal rhetoric remains minimally consistent with observable political reality, blame should be more common around opposing-party economic legislative action, while credit should be more common around own-party economic legislative action. The same pattern would also make the claim politically convenient: blame would be attached to the opposing party, while credit would be attached to one's own party.

To examine this idea, we construct bill-exposure indicators at the date-by-chamber-by-party level, separately for the House and Senate. For each politician-day, we define OwnBill_{ict} as an indicator equal to one if, on date t , at least one economic bill is introduced in politician i 's chamber c by a member of politician i 's party. We define OppBill_{ict} analogously for economic bills introduced in the same chamber by members of the opposing party.

In this exercise, we restrict attention to the period in which both House and Senate tweets are observed and include all 537,998 tweets identified by our topic classifier as being about the economy, whether or not they contain causal rhetoric. At the tweet level, let $y_{\ell ct}$ be an indicator equal to one if economic tweet ℓ , posted by politician i in chamber c on date t , is classified as rhetorical style y , where $y \in \{\text{Blame}, \text{Credit}\}$. We

Table 4: Economic Causal Rhetoric around Economic Bill Introductions

	(1)	(2)
	Blame	Credit
Bill introduced by opposing party	0.014** (0.007)	-0.002 (0.006)
Bill introduced by own party	-0.006 (0.006)	0.018** (0.007)
Politician x Month FE	✓	✓
Date x Party FE	✓	✓
Observations	535,776	535,776
Clusters	2,203	2,203

Notes: The table presents estimates of Equation 3. The sample includes all tweets classified as being about the economy. The outcome in column (1) is an indicator equal to one if the tweet is classified as blame. The outcome in column (2) is an indicator equal to one if the tweet is classified as credit. The regressors indicate whether, on the date of the tweet, at least one economic bill was introduced in the politician’s chamber by a member of the opposing party or by a member of the politician’s own party. All specifications include politician-by-month and date-by-politician-party fixed effects. Standard errors in parentheses are clustered by date. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

estimate separate specifications for blame and credit:

$$y_{lict} = \beta_{Opp} OppBill_{ict} + \beta_{Own} OwnBill_{ict} + \lambda_{i \times m(t)} + \mu_{t \times P_i} + \varepsilon_{lict}. \quad (3)$$

The fixed effects $\lambda_{i \times m(t)}$ are politician-by-month fixed effects, which absorb politician-specific differences in rhetorical style and month-specific changes in each politician’s communication. The fixed effects $\mu_{t \times P_i}$ are date-by-party fixed effects, where P_i denotes politician i ’s party. They absorb shocks common to politicians from the same party on the same date. Standard errors are clustered by date.

The coefficients describe conditional differences in economic rhetorical style around chamber-specific economic legislative activity. Date-by-party fixed effects compare politicians from the same party on the same date, while politician-by-month fixed effects compare each politician to themselves within the same month. The remaining variation comes from differences in economic bill introductions across the House and Senate within the same party and date. Because bill introductions are political events rather than randomly assigned shocks, we interpret the estimates descriptively.

Table 4 reports estimates of Equation 3. Economic causal rhetoric moves with legislative activity in a way that is both plausible and politically aligned. On dates when the opposing party introduces an economic bill in a politician’s chamber, economic tweets by politicians in that chamber are 1.4 percentage points more likely to be classified as blame, relative to economic tweets by same-party politicians on the same date; own-party economic bill introductions are not associated with a comparable increase in blame. For credit, the pattern is reversed: own-party economic bill

introductions are associated with a 1.8 percentage point higher probability that an economic tweet is classified as credit, while opposing-party economic bill introductions are not associated with a comparable increase. Thus, economic blame is more common around opposing-party legislative action, while economic credit is more common around own-party legislative action.

The bill evidence adds the plausibility and political-convenience sides of the interpretation. It reinforces plausibility because economic causal rhetoric lines up with observable political actions that make the corresponding claim more plausible. But the alignment is not neutral. Blame rises around economic actions taken by the opposing party, while credit rises around economic actions taken by one's own party. These patterns suggest that economic causal rhetoric remains plausible by staying minimally consistent with political reality, while assigning blame and credit in a politically convenient direction.

Taken together, the two exercises help to characterize the claims contained in economic causal rhetoric. The content analysis shows that these claims are usually underqualified, but not usually outlandish. The bill analysis shows that they are used around observable political actions that make the corresponding claim plausible, and that this use is politically aligned: blame attaches to opposing-party action, while credit attaches to own-party action. This distinction is most important for blame, the rhetorical style whose political returns are largest and whose economic claims are most likely to require substantial qualifications. The evidence points to a subtle margin: causal rhetoric, and especially blame, can turn complex economic mechanisms into underqualified yet plausible, politically convenient claims.

7 Conclusion

Democratic politics asks voters to form beliefs about how political actions map into outcomes. This paper studies a central rhetorical margin in that process: politicians' use of language that links the actual or potential action of a political actor to an outcome through blame and credit. We call this language causal rhetoric. We measure its use in congressional tweets by fine-tuning a language model. The resulting evidence shows that causal rhetoric is central to political communication. It accounts for more than one third of congressional tweets, is the language of political disagreement, and has become substantially more prevalent over time.

The political returns to causal rhetoric are asymmetric and concentrated in blame. Exposure to blame generates sizable and consistent gains in political support, while credit produces no comparable aggregate response. In campaign-contributions data,

blame increases small-donation revenue primarily by mobilizing additional donors, while credit produces smaller returns detectable only in average giving. Survey evidence for the 2020 House elections points in the same direction: support rises after blame-only days, especially among respondents who use social media for political news. Descriptive evidence on retweets suggests one possible reason for this asymmetry: blame travels farther than credit by receiving more retweets.

The evidence on economic causal rhetoric clarifies what kind of claims it contains. The content analysis shows that these claims are usually underqualified, but not usually outlandish: they often require qualifications that the tweet omits, while outright unsupported claims are rare. The bill analysis shows that these claims also line up with observable political actions in a politically convenient direction: economic blame rises around opposing-party legislative action, while economic credit rises around own-party legislative action. Causal rhetoric therefore occupies a subtle margin. It can turn complex economic mechanisms into underqualified yet plausible, politically convenient claims, especially in blame, where political returns are concentrated.

The findings point to two directions for future research. A first is to identify why blame is more effective than credit. Behavioral work, especially using survey and experimental designs, could distinguish between mechanisms such as attention, memory, emotional activation, and social transmission. A second direction is to assess the welfare implications of causal rhetoric. Fine-grained individual-level data could help measure whether these claims improve voters' understanding of political responsibility or instead distort beliefs about how political actions generate outcomes.

Overall, the paper shows that causal rhetoric is a major input into democratic politics. Politicians frequently tell voters how political actions and outcomes are connected, and these claims generate political returns, especially when they assign blame. Understanding democratic accountability therefore requires studying not only how voters evaluate outcomes, but also how politicians shape the causal accounts through which those outcomes are understood.

Bibliography

- Achen, C. H. and L. M. Bartels (2016). *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.
- Algan, Y., E. Davoine, T. Renault, and S. Stantcheva (2026). Emotions and policy views. *Working Paper*.
- Andre, P., I. Haaland, C. Roth, M. Wiederholt, and J. Wohlfart (2026). Narratives about the macroeconomy. *The Review of Economic Studies*, rdag014.
- Aridor, G., R. Jiménez-Durán, R. Levy, and L. Song (2024). The economics of social media. *Journal of Economic Literature* 62(4), 1422–1474.
- Ashworth, S. (2012). Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science* 15(1), 183–201.
- Bellodi, L., M. Morelli, A. Nicolo, and P. Roberti (2025). The shift to commitment politics and populism: Theory and evidence. *BAFFI CAREFIN Working Paper n. 715*.
- Besley, T. (2006). *Principled agents?: The political economy of good government*. Oxford University Press.
- Besley, T. and A. Case (1995). Does electoral accountability affect economic policy choices? evidence from gubernatorial term limits. *The Quarterly Journal of Economics* 110(3), 769–798.
- Bilotta, F. and G. Manferdini (2026). Coarse memory and plausible narratives. *American Economic Journal: Microeconomics*. Forthcoming.
- Boken, J., M. Draca, N. Mastrococco, and A. Ornaghi (2023). The returns to viral media: the case of us campaign contributions. *Working Paper*.
- 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. <https://data.stanford.edu/dime>.
- Borusyak, K., P. Hull, and X. Jaravel (2025). A practical guide to shift-share instruments. *Journal of Economic Perspectives* 39(1), 181–204.
- Bouton, L., J. Cagé, E. Dewitte, and V. Pons (2022). Small campaign donors. *NBER Working Paper No. 30050*.

- Burnham, M., K. Kahn, R. Y. Wang, and R. X. Peng (2025). Political debate: Efficient zero-shot and few-shot classifiers for political text. *Political Analysis*, 1–15.
- 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*.
- Canes-Wrone, B., M. C. Herron, and K. W. Shotts (2001). Leadership and pandering: A theory of executive policymaking. *American Journal of Political Science*, 532–550.
- Chater, N. and G. Loewenstein (2016). The under-appreciated drive for sense-making. *Journal of Economic Behavior & Organization* 126, 137–154.
- Chen, J. and J. Roth (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics* 139(2), 891–936.
- Davidson, T., D. Warmusley, 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. *Annual Review of Economics* 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.
- Eliaz, K., S. Galperti, and R. Spiegler (2025). False narratives and political mobilization. *Journal of the European Economic Association* 23(3), 983–1027.
- Eliaz, K. and A. Rubinstein (2025). Wasonian persuasion.
- Eliaz, K. and R. Spiegler (2020). A model of competing narratives. *American Economic Review* 110(12), 3786–3816.
- Enke, B. (2020). Moral values and voting. *Journal of Political Economy* 128(10), 3679–3729.

- Federal Election Commission (2020). 2020 presidential and congressional primary dates and candidate filing deadlines for ballot access. <https://www.fec.gov/resources/cms-content/documents/2020pdates.pdf>. Data as of July 17, 2020. Accessed July 1, 2026.
- Ferraz, C. and F. Finan (2008). Exposing corrupt politicians: the effects of brazil’s publicly released audits on electoral outcomes. *The Quarterly Journal of Economics* 123(2), 703–745.
- Ferraz, C. and F. Finan (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review* 101(4), 1274–1311.
- 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 (2025). Virality: What makes narratives go viral, and does it matter? *Working Paper*.
- Gennaro, G. and E. Ash (2022). Emotion and reason in political language. *The Economic Journal* 132(643), 1037–1059.
- 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 (2024). Crash narratives. *NBER Working Paper No. 30195*.
- Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik instruments: What, when, why, and how. *American Economic Review* 110(8), 2586–2624.
- Graeber, T., C. Roth, and C. Schesch (2026). Explanations. *University of Zurich Working Paper No. 490*.
- He, H. and E. A. Garcia (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering* 21(9), 1263–1284.
- Healy, A. and N. Malhotra (2013). Retrospective voting reconsidered. *Annual Review of Political Science* 16, 285–306.
- Hüning, H., L. Mechtenberg, and S. Wang (2022). Using arguments to persuade: Experimental evidence. *SSRN Working Paper*.

- 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 (2024). Narrative-driven fluctuations in sentiment: Evidence linking traditional and social media. *Working Paper*.
- Malle, B. F., S. Guglielmo, and A. E. Monroe (2014). A theory of blame. *Psychological Inquiry* 25(2), 147–186.
- Maskin, E. and J. Tirole (2004). The politician and the judge: Accountability in government. *American Economic Review* 94(4), 1034–1054.
- 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*.
- 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.
- 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. *Working Paper*.
- Schmitter, P. C. and T. L. Karl (1991). What democracy is... and is not. *Journal of democracy* 2(3), 75–88.
- Sloman, S. A. and D. Lagnado (2015). Causality in thought. *Annual Review of Psychology* 66, 223–247.
- Snyder Jr, J. M. and D. Strömberg (2010). Press coverage and political accountability. *Journal of Political Economy* 118(2), 355–408.

- 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.
- Tausanovitch, C. and L. Vavreck (2021). Democracy fund + ucla nationscape project.
- Trump Twitter Archive (2021). <https://www.thetrumparchive.com>.
- Tweets of Congress (2023). <https://github.com/alexlitel/congresstweets>.
- Westwood, S. and Y. Lelkes (2024). America’s political pulse [data set].
- 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 Figures and Tables

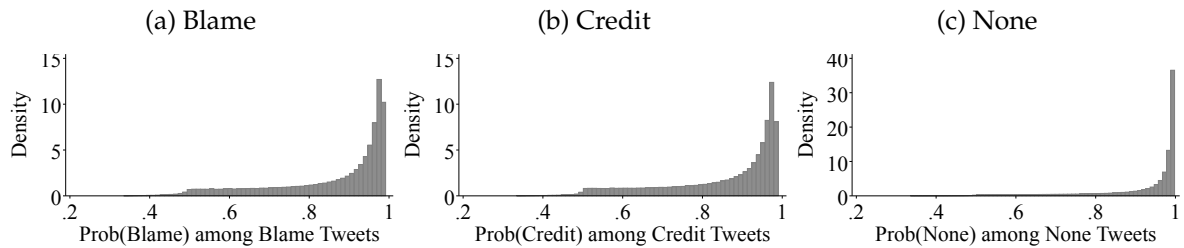
A.1 Data and Measurement

Table A1: Descriptive Statistics

	(a) Tweet Level			(b) Politician Level		
	All	Democrats	Republicans	All	Democrats	Republicans
Female	0.300 (0.458)	0.401 (0.490)	0.165 (0.371)	0.243 (0.429)	0.378 (0.486)	0.132 (0.339)
Age	57.973 (11.572)	59.058 (12.075)	56.509 (10.682)	57.468 (11.767)	58.499 (12.649)	56.617 (10.926)
Black	0.082 (0.274)	0.127 (0.333)	0.021 (0.142)	0.080 (0.271)	0.150 (0.357)	0.022 (0.148)
Bachelor	0.333 (0.471)	0.297 (0.457)	0.381 (0.486)	0.332 (0.471)	0.295 (0.457)	0.363 (0.481)
Master or Higher	0.620 (0.485)	0.672 (0.469)	0.551 (0.497)	0.584 (0.493)	0.651 (0.477)	0.529 (0.500)
Republican	0.426 (0.494)	0 (0)	1 (0)	0.548 (0.498)	0 (0)	1 (0)
Nominate Score	-0.005 (0.467)	-0.387 (0.126)	0.511 (0.164)	0.109 (0.455)	-0.366 (0.122)	0.500 (0.162)
Representative	0.804 (0.397)	0.814 (0.389)	0.790 (0.407)	0.857 (0.351)	0.857 (0.350)	0.856 (0.351)
Senator				0.118 (0.323)	0.123 (0.329)	0.114 (0.318)
Both Representative and Senator				0.026 (0.158)	0.020 (0.139)	0.030 (0.172)
Number of Tweets Posted				4664.950 (4848.342)	5924.391 (5151.747)	3625.209 (4319.150)
Number of Accounts				1.988 (0.689)	2.042 (0.688)	1.943 (0.687)
Observations	4,198,455	2,411,227	1,787,228	900	407	493

Notes: Panel (a) presents statistics at the tweet level. Panel (b) presents statistics at the politician level. Means and standard deviations in parentheses.

Figure A1: Distributions



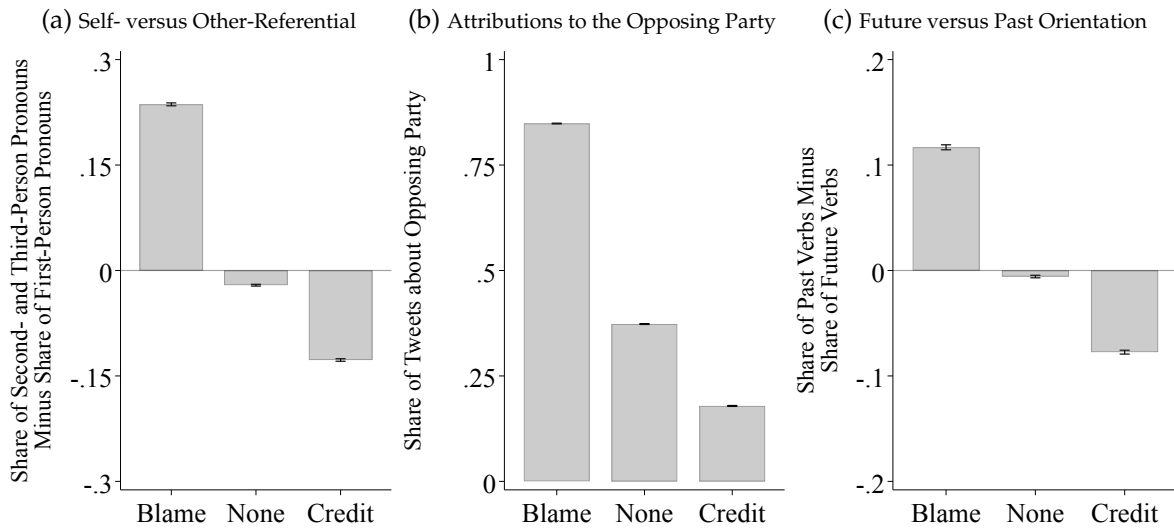
Notes: Each panel reports the distribution of the classifier-assigned probability for the predicted class, conditional on that class being selected. Panel (a) shows the distribution of $P(\text{Blame})$ among tweets classified as blame, that is, tweets for which $P(\text{Blame}) > \max\{P(\text{Credit}), P(\text{None})\}$. Panels (b) and (c) report the analogous distributions for credit and none.

Table A2: Labeled Tweets

Tweet	Causal	Tone	Rhetoric Style
Biden has lost all credibility. No one believes your lies, Joe! URL	0	-1	None
.realDonaldTrump is a terrible role model for our kids. They deserve better. #Debate	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	Credit
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	Credit

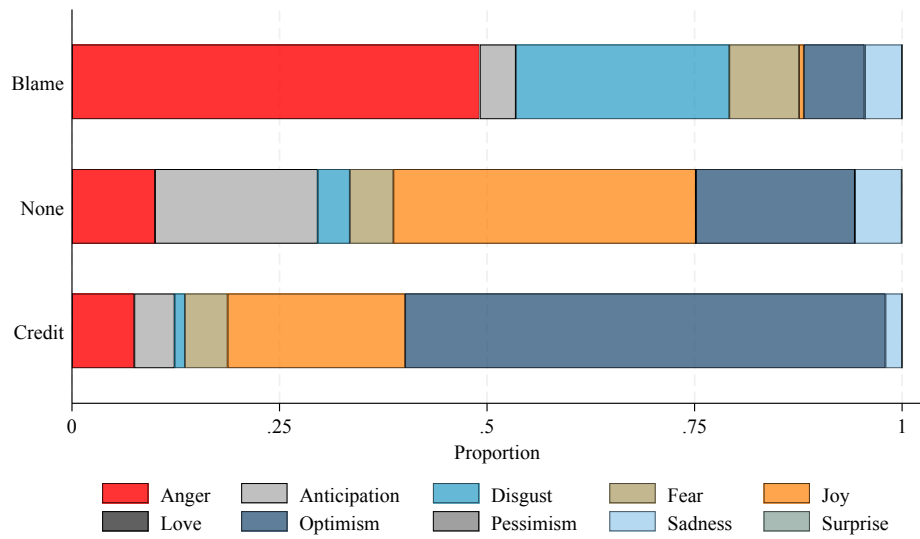
Notes: The table reports illustrative tweets from the hand-labeled dataset used to train the classifier. For each tweet, we report the manual labels for causality and tone, together with the implied rhetorical-style classification.

Figure A2: Linguistic Features



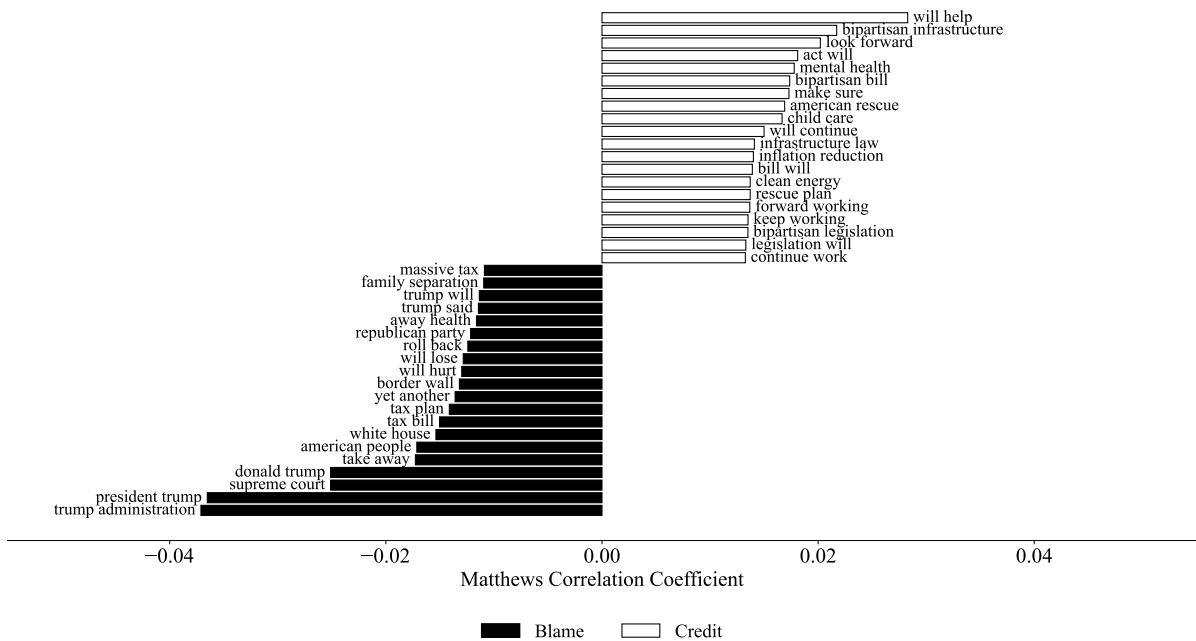
Notes: The figure reports linguistic and semantic measures across rhetorical styles. Panel (a) compares the relative prevalence of second- and third-person pronouns to first-person pronouns, capturing whether tweets are more other- or self-referential. Panel (b) reports the share of tweets that target the opposing party, among tweets for which party identity can be identified. Panel (c) compares the use of future and past tense, capturing whether the rhetoric is more prospective or retrospective. The measures in panels (a) and (c) are standardized. Bars represent 95 percent confidence intervals.

Figure A3: Emotions



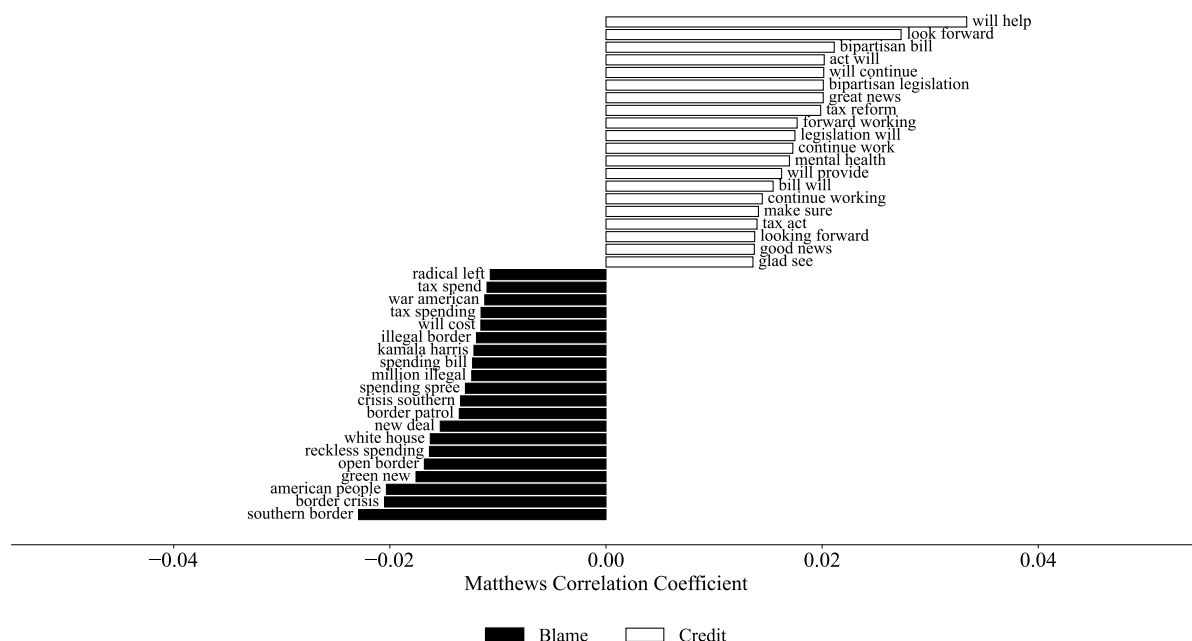
Notes: Each bar reports the share of tweets within a given rhetorical style classified into the corresponding emotion category.

Figure A4: Bigrams Diagnostic of Blame and Credit for Democrats



Notes: The figure reports the 20 bigrams that are most distinctive of blame and credit tweets among Democratic politicians, ranked by their Matthews correlation coefficient.

Figure A5: Bigrams Diagnostic of Blame and Credit for Republicans



Notes: The figure reports the 20 bigrams that are most distinctive of blame and credit tweets among Republican politicians, ranked by their Matthews correlation coefficient.

Table A3: Democrats Tweets Classified as Blame and Credit

(a) 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.	
(b) Credit 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.	

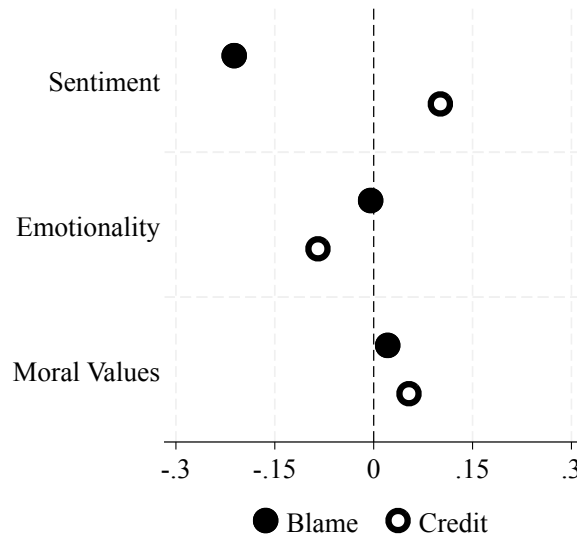
Notes: The table reports illustrative examples of tweets classified as blame and credit among Democratic politicians. Panel (a) reports blame tweets and Panel (b) reports credit tweets. These examples are intended to provide qualitative intuition for the rhetorical patterns captured by our classification.

Table A4: Republicans Tweets Classified as Blame and Credit

(a) 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	
(b) Credit 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	

Notes: The table reports illustrative examples of tweets classified as blame and credit among Republican politicians. Panel (a) reports blame tweets and Panel (b) reports credit tweets. These examples are intended to provide qualitative intuition for the rhetorical patterns captured by our classification.

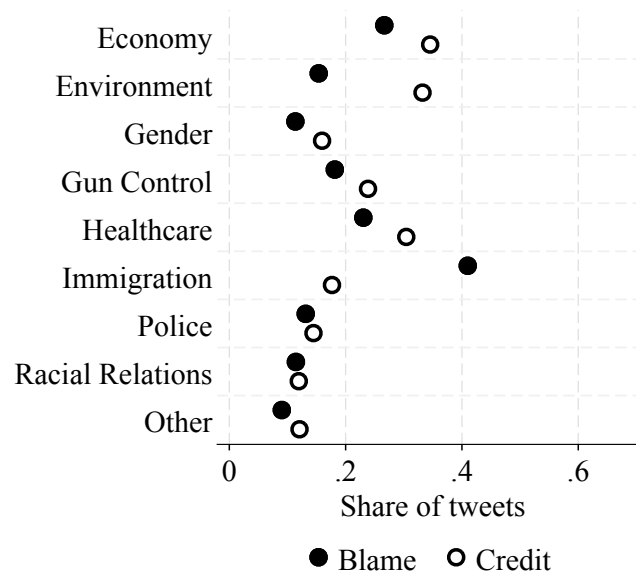
Figure A6: Distinctiveness Relative to Existing Text Measures



Notes: The figure reports coefficients from tweet-level regressions in which the outcome is an indicator equal to one if the tweet is classified as blame or credit. Each outcome is regressed on three standardized text measures: sentiment, emotionality, and moral terminology. Sentiment is measured using the [Camacho-Collados et al. \(2022\)](#) model, which assigns each tweet probabilities of being negative, neutral, or positive; we convert these probabilities into a continuous sentiment measure equal to $-1 \times \text{Pr}(\text{negative}) + 0 \times \text{Pr}(\text{neutral}) + 1 \times \text{Pr}(\text{positive})$. Emotionality follows [Gennaro and Ash \(2022\)](#): we embed emotional and reasoning words and compute, for each tweet, the relative cosine similarity of the tweet embedding to the emotional-word embedding versus the reasoning-word embedding. Moral terminology follows [Enke \(2020\)](#): for each moral value, we compute the average frequency of vice and virtue terms, sum across moral values, and normalize by the number of non-stop words. All regressors are standardized. Bars represent 95 percent confidence intervals computed with standard errors clustered at the politician level.

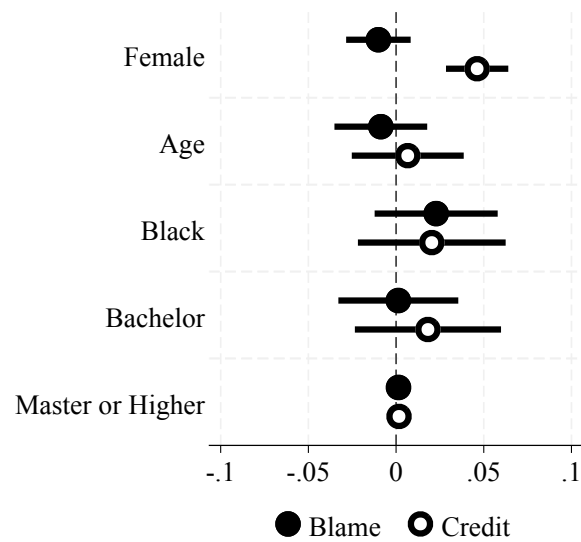
A.2 Supply

Figure A7: Blame and Credit over Topics



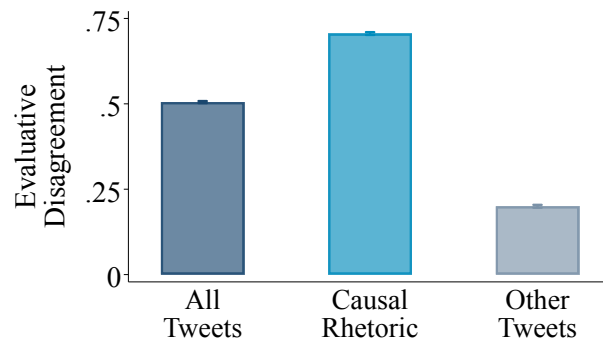
Notes: The figure presents the share of blame and credit tweets within each topic. Bars represent 95 percent confidence intervals.

Figure A8: Author Correlates of Blame and Credit



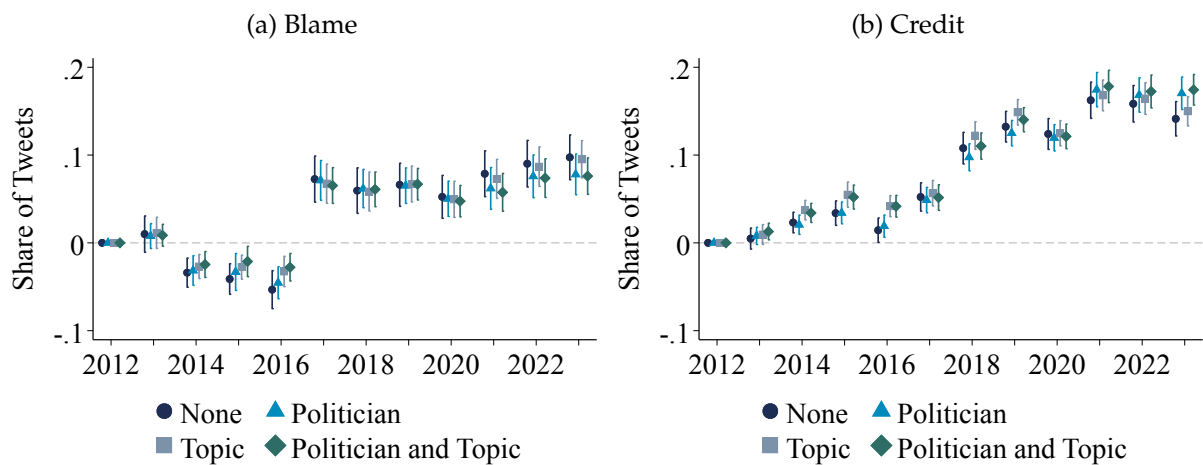
Notes: The figure present the estimates from two regressions at the politician level of blame and credit over the listed variables. Bars represent 95 percent confidence intervals computed with robust standard errors.

Figure A9: Causal Rhetoric and Disagreement among Targeted Tweets



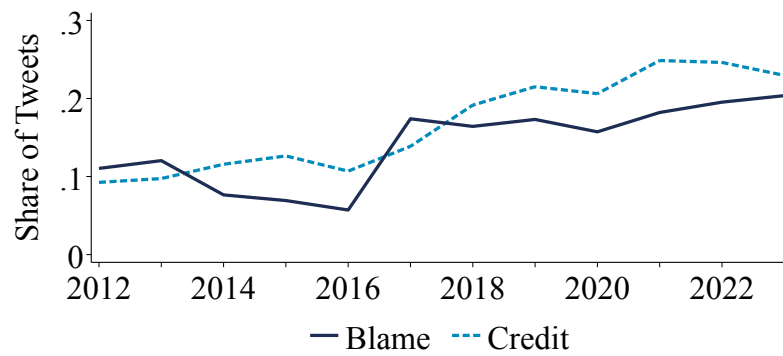
Notes: The figure reports evaluative disagreement among targeted tweets, measured as the difference in average standardized sentiment between tweets posted by members of the ruling party and tweets posted by members of the opposition. The sample is restricted to tweets that can be identified as targeting either Democrats or Republicans. Higher values indicate a larger divergence in how politicians on opposite sides of political control evaluate the political environment. We compute this measure separately for all targeted tweets, targeted tweets classified as blame or credit, and targeted tweets classified as neither blame nor credit. Bars represent 95 percent confidence intervals.

Figure A10: Blame and Credit over Time: Adjusting for Politician and Topic Composition



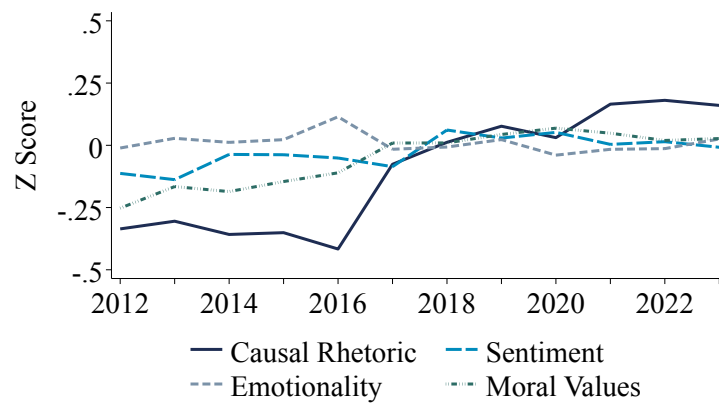
Notes: The figure reports yearly changes in the share of tweets classified as blame or credit, relative to 2012. Each series comes from a specification that progressively adds the fixed effects indicated in the legend. Politician fixed effects absorb changes in the composition of members of Congress in the sample. Topic fixed effects absorb changes in the issues politicians discuss. Bars represent 95 percent confidence intervals computed with standard errors clustered at the politician level.

Figure A11: Blame and Credit over Time: Excluding Senators



Notes: The figure reports the yearly share of tweets classified as blame and credit after excluding tweets posted by Senators.

Figure A12: Comparison with Other Text Measures over Time



Notes: The figure reports yearly averages of causal rhetoric, sentiment, emotionality, and moral language. All measures are standardized.

A.3 Political Returns

Table A5: List of Underlying County-Level Covariates

Variable
Population decile
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 lists the 29 county-level covariates used to construct the principal components that are interacted with blame, credit, and tweet volume in Equations 1.

Table A6: Donation Returns: First Stage

	(1)	(2)
	Blame x Users	Credit x Users
Blame x SXSWFollower2007	2.482*** (0.180)	-0.000*** (0.000)
Credit x SXSWFollower2007	-0.000*** (0.000)	2.482*** (0.180)
Blame x SXSWFollower2006	-0.274 (0.394)	0.000 (0.000)
Credit x SXSWFollower2006	0.000 (0.000)	-0.274 (0.394)
Politician x Month FE	✓	✓
Politician x County FE	✓	✓
County x Month FE	✓	✓
Tweet Volume Control	✓	✓
Observations	168,232,932	168,232,932
Clusters	3,108	3,108

Notes: The table presents first-stage estimates for the preferred 2SLS specification in Equation 1. The dependent variable is the interaction between the standardized monthly share of blame tweets and county Twitter users in column (1), and the interaction between the standardized monthly share of credit tweets and county Twitter users in column (2). The main excluded instruments are the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2007. The table also reports the coefficients on the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2006. All specifications include politician-by-month, politician-by-county, and county-by-month fixed effects, as well as the interaction between tweet volume and county Twitter users. Standard errors in parentheses are clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A7: Donation Returns: OLS Estimates

	(1)	(2)	(3)
	Revenue from Donations	Number of Donors	Average per Donor
Blame x Users	0.006*** (0.000)	0.002*** (0.000)	0.004*** (0.000)
Credit x Users	-0.000 (0.000)	-0.000 (0.000)	-0.000** (0.000)
Politician x Month FE	✓	✓	✓
Politician x County FE	✓	✓	✓
County x Month FE	✓	✓	✓
Tweet Volume Control	✓	✓	✓
Observations	168,232,932	168,232,932	168,232,932
Clusters	3,108	3,108	3,108

Notes: The table presents OLS estimates of Equation 1. The outcomes are log-donation revenue in column (1), log-number of donors in column (2), and log-average donation per donor in column (3). The main regressors are the standardized monthly shares of blame and credit tweets posted by the politician, each interacted with the log of one plus the number of Twitter users in the county. All specifications include politician-by-month, politician-by-county, and county-by-month fixed effects, as well as the interaction between tweet volume and county Twitter users. Standard errors in parentheses are clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A8: Donation Returns: Reduced Form Estimates

	(1)	(2)	(3)
	Revenue from Donations	Number of Donors	Average per Donor
Blame x SXSWFollower2007	0.013*** (0.003)	0.006*** (0.001)	0.008*** (0.002)
Credit x SXSWFollower2007	0.003*** (0.001)	0.001** (0.000)	0.002*** (0.001)
Blame x SXSWFollower2006	-0.008 (0.005)	-0.003 (0.002)	-0.005 (0.003)
Credit x SXSWFollower2006	0.003 (0.002)	0.001 (0.001)	0.002 (0.001)
Politician x Month FE	✓	✓	✓
Politician x County FE	✓	✓	✓
County x Month FE	✓	✓	✓
Tweet Volume Control	✓	✓	✓
Observations	168,232,932	168,232,932	168,232,932
Clusters	3,108	3,108	3,108

Notes: The table presents reduced-form estimates corresponding to Equation 1. The outcomes are log-donation revenue in column (1), log-number of donors in column (2), and log-average donation per donor in column (3). The main regressors are the standardized monthly shares of blame and credit tweets posted by the politician, each interacted with the log of one plus the number of SXSW followers in the county who joined Twitter in 2007. The table also reports the coefficients on the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2006. All specifications include politician-by-month, politician-by-county, and county-by-month fixed effects, as well as the interaction between tweet volume and county Twitter users. Standard errors in parentheses are clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A9: Donation Returns: Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Revenue from Donations			Number of Donors			Average per Donor		
Blame x Users	0.006*** (0.001)	0.006*** (0.001)	0.006** (0.003)	0.003*** (0.001)	0.003*** (0.000)	0.004*** (0.001)	0.003*** (0.001)	0.003*** (0.001)	0.002 (0.002)
Credit x Users	0.002*** (0.000)	0.002*** (0.000)	0.002** (0.001)	0.001*** (0.000)	0.001*** (0.000)	0.000 (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.002** (0.001)
Blame x SXSWFollower2006	-0.006 (0.004)	-0.006 (0.004)	-0.003 (0.004)	-0.003 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.004 (0.003)	-0.004 (0.003)	-0.002 (0.003)
Credit x SXSWFollower2006	0.003 (0.002)	0.002 (0.002)	0.000 (0.002)	0.001 (0.001)	0.001 (0.001)	-0.000 (0.001)	0.002* (0.001)	0.002 (0.001)	0.000 (0.001)
Politician x Month FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tweet Volume Control	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sentiment Control	✓			✓			✓		
Attack/Praise Control		✓			✓			✓	
County Covariates Controls			✓			✓			✓
Observations	168,232,932	168,232,932	168,178,803	168,232,932	168,232,932	168,178,803	168,232,932	168,232,932	168,178,803
Clusters	3,108	3,108	3,107	3,108	3,108	3,107	3,108	3,108	3,107
F statistic	38.19	38.19	20.20	38.19	38.19	20.20	38.19	38.19	20.20

Notes: The table presents 2SLS estimates of Equation 1 under alternative control specifications. Columns (1)–(3) use log-donation revenue as the outcome, columns (4)–(6) use log-number of donors, and columns (7)–(9) use log-average donation per donor. Within each outcome, the first column adds controls for exposure to non-causal negative and positive sentiment, the second column adds controls for exposure to non-causal attack and praise, and the third column allows the effects of blame, credit, and tweet volume to vary with county covariates. The main regressors are the standardized monthly shares of blame and credit tweets posted by the politician, each interacted with the log of one plus the number of Twitter users in the county. The table also reports the coefficients on the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2006. Interactions with county Twitter users are instrumented using the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2007. All specifications include politician-by-month, politician-by-county, and county-by-month fixed effects, as well as the interaction between tweet volume and county Twitter users. Standard errors in parentheses are clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A10: Donation Returns: Any Donation

	(1)	(2)	(3)	(4)
Blame x Users	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.002** (0.001)
Credit x Users	-0.000 (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Blame x SXSWFollower2006	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.000 (0.001)
Credit x SXSWFollower2006	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Politician x Month FE	✓	✓	✓	✓
Politician x County FE	✓	✓	✓	✓
County x Month FE	✓	✓	✓	✓
Tweet Volume Control	✓	✓	✓	✓
Sentiment Control		✓		
Attack/Praise Control			✓	
County Covariates Controls				✓
Observations	168,232,932	168,232,932	168,232,932	168,178,803
Clusters	3,108	3,108	3,108	3,107
F statistic	63.65	38.19	38.19	20.20

Notes: The table presents 2SLS estimates of Equation 1 using as outcome an indicator equal to one if the politician receives at least one donation from the county in the month. Column (1) reports the baseline specification with the tweet-volume control. Column (2) adds controls for exposure to non-causal negative and positive sentiment. Column (3) adds controls for exposure to non-causal attack and praise. Column (4) allows the effects of blame, credit, and tweet volume to vary with county covariates. The main regressors are the standardized monthly shares of blame and credit tweets posted by the politician, each interacted with the log of one plus the number of Twitter users in the county. The table also reports the coefficients on the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2006. Interactions with county Twitter users are instrumented using the corresponding interactions with the log of one plus the number of SXSW followers in the county who joined Twitter in 2007. All specifications include politician-by-month, politician-by-county, and county-by-month fixed effects, as well as the interaction between tweet volume and county Twitter users. Standard errors in parentheses are clustered at the county level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A11: Balance in Respondents' Observable Characteristics

	(1)	(2)	(3)	(4)	(5)
	Female	Age	White	College	Employed
(a) Blame					
Post-Blame	-0.008 (0.008)	-0.200 (0.255)	-0.011* (0.007)	-0.005 (0.008)	-0.006 (0.008)
District x Month FE	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓
Partisan Alignment FE	✓	✓	✓	✓	✓
Observations	20,154	20,154	20,154	20,154	20,154
Clusters	281	281	281	281	281
(b) Credit					
Post-Credit	-0.002 (0.006)	-0.305 (0.225)	-0.003 (0.005)	-0.001 (0.006)	0.008 (0.007)
District x Month FE	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓
Partisan Alignment FE	✓	✓	✓	✓	✓
Observations	29,109	29,109	29,109	29,109	29,109
Clusters	325	325	325	325	325

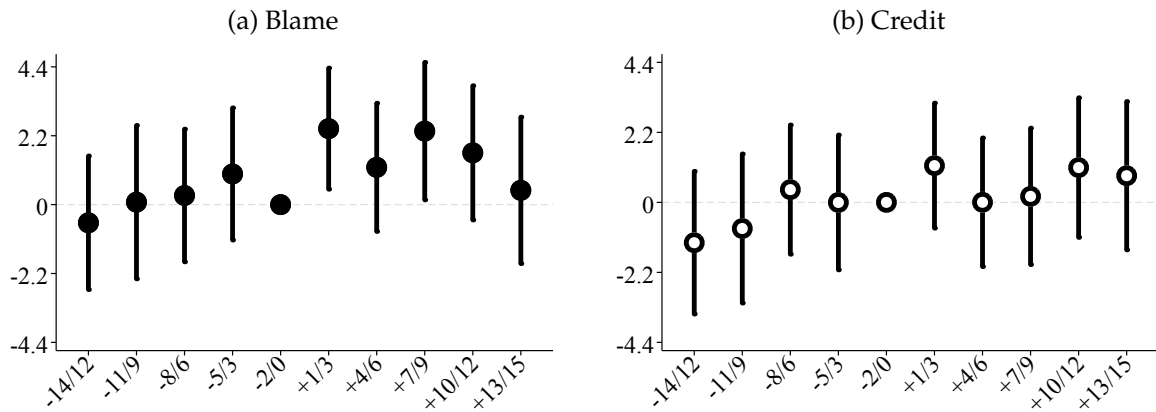
Notes: The table presents balance checks for respondents' observable characteristics around rhetoric-only days. Each coefficient comes from a separate regression of the column variable on the post-period indicator. Panel (a) reports estimates around blame-only days, and Panel (b) reports estimates around credit-only days. The treatment indicator equals one for respondents interviewed in the seven days after a rhetoric-only day. The outcomes are an indicator for being female, age, an indicator for being white, an indicator for having a college degree, and an indicator for being employed. All regressions include district-by-month, date, and partisan alignment fixed effects. Standard errors in parentheses are clustered at the congressional-district level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A12: Balance in Respondent–Politician Partisan Alignment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dem-Dem	Rep-Dem	Ind-Dem	Else-Dem	Dem-Rep	Rep-Rep	Ind-Rep	Else-Rep
(a) Blame								
Post-Blame	-0.010 (0.007)	0.006 (0.007)	0.005 (0.005)	-0.000 (0.003)	-0.003 (0.004)	0.006 (0.004)	-0.004 (0.003)	0.000 (0.002)
District x Month FE	✓	✓	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓	✓	✓
Respondent Covariate FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	20,154	20,154	20,154	20,154	20,154	20,154	20,154	20,154
Clusters	281	281	281	281	281	281	281	281
(b) Credit								
Post-Credit	-0.000 (0.005)	0.004 (0.004)	-0.007* (0.004)	0.003 (0.002)	-0.004 (0.004)	0.001 (0.005)	0.005 (0.004)	-0.002 (0.002)
District x Month FE	✓	✓	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓	✓	✓
Respondent Covariate FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	29,109	29,109	29,109	29,109	29,109	29,109	29,109	29,109
Clusters	325	325	325	325	325	325	325	325

Notes: The table presents balance checks for respondent–politician partisan alignment around rhetoric-only days. Each coefficient comes from a separate regression of the column variable on the post-period indicator. Panel (a) reports estimates around blame-only days, and Panel (b) reports estimates around credit-only days. The treatment indicator equals one for respondents interviewed in the seven days after a rhetoric-only day. The outcomes are indicators for the interaction between respondent party identification and the politician’s party. The first term in each column label denotes the respondent’s party identification; the second term denotes the politician’s party. All regressions include district-by-month, date, and respondent covariate fixed effects. Standard errors in parentheses are clustered at the congressional-district level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Figure A13: Voting Intentions: Dynamics



Notes: The figure presents dynamic estimates around rhetoric-only days. Panel (a) reports estimates around blame-only days, and Panel (b) reports estimates around credit-only days. The omitted category is the rhetoric-only day and the two days immediately before it, corresponding to days -2 to 0. The remaining coefficients are three-day event-time bins before and after the omitted period. The specification includes the fixed effects in Equation 2. Observations that fall into more than one event-time bin are excluded. Bars represent 95 percent confidence intervals computed with standard errors clustered at the congressional-district level.

Table A13: Voting Intentions: Date Fixed-Effect Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
	All Respondents		Split by Social Media Use			
			No Social Media Use		Social Media Use	
	(a) Blame					
Post-Blame	0.011** (0.005)	0.009* (0.005)	0.002 (0.011)	0.007 (0.011)	0.015** (0.006)	0.013** (0.006)
District x Month FE	✓	✓	✓	✓	✓	✓
Partisan Alignment FE	✓	✓	✓	✓	✓	✓
Respondent Covariate FE	✓	✓	✓	✓	✓	✓
Date x Politician’s Party FE	✓		✓		✓	
Date x Respondent’s Party FE		✓		✓		✓
Observations	20,144	20,105	5,757	5,646	14,284	14,227
Clusters	281	281	275	275	281	281
	(b) Credit					
Post-Credit	0.006 (0.004)	0.007 (0.004)	0.003 (0.010)	-0.001 (0.009)	0.005 (0.005)	0.007 (0.005)
District x Month FE	✓	✓	✓	✓	✓	✓
Partisan Alignment FE	✓	✓	✓	✓	✓	✓
Respondent Covariate FE	✓	✓	✓	✓	✓	✓
Date x Politician’s Party FE	✓		✓		✓	
Date x Respondent’s Party FE		✓		✓		✓
Observations	29,101	29,080	8,239	8,178	20,746	20,716
Clusters	325	325	322	322	325	325

Notes: The table presents estimates of Equation 2 under alternative date fixed-effect specifications. The outcome is an indicator equal to one if the respondent reports intending to vote for the politician, and zero otherwise. Panel (a) reports estimates for blame-only days, and Panel (b) reports estimates for credit-only days. The treatment indicator equals one for respondents interviewed in the seven days after a blame-only or credit-only day. Columns (1) and (2) use all respondents. Columns (3) and (4) restrict the sample to respondents who do not report having seen or heard political news on social media in the past week. Columns (5) and (6) restrict the sample to respondents who do report having seen or heard political news on social media in the past week. Columns (1), (3), and (5) include date-by-politician-party fixed effects, while columns (2), (4), and (6) include date-by-respondent-party fixed effects. All specifications include district-by-month, partisan alignment, and respondent covariate fixed effects. Standard errors in parentheses are clustered at the congressional-district level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A14: Voting Intentions: Tweet-Volume Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
	All Respondents		Split by Social Media Use			
			No Social Media Use		Social Media Use	
	(a) Blame					
Post-Blame	0.010** (0.005)	0.010** (0.005)	0.004 (0.010)	0.006 (0.011)	0.014** (0.006)	0.014** (0.006)
District x Month FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Partisan Alignment FE	✓	✓	✓	✓	✓	✓
Respondent Covariate FE	✓	✓	✓	✓	✓	✓
Rhetoric-Day Volume FE	✓		✓		✓	
Window Volume Control		✓		✓		✓
Observations	20,154	20,154	5,781	5,781	14,300	14,300
Clusters	281	281	275	275	281	281
	(b) Credit					
Post-Credit	0.006 (0.004)	0.006 (0.004)	-0.001 (0.009)	-0.000 (0.009)	0.006 (0.005)	0.006 (0.005)
District x Month FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Partisan Alignment FE	✓	✓	✓	✓	✓	✓
Respondent Covariate FE	✓	✓	✓	✓	✓	✓
Rhetoric-Day Volume FE	✓		✓		✓	
Window Volume Control		✓		✓		✓
Observations	29,109	29,109	8,257	8,257	20,756	20,756
Clusters	325	325	322	322	325	325

Notes: The table presents estimates of Equation 2 under alternative date fixed-effect specifications. The outcome is an indicator equal to one if the respondent reports intending to vote for the politician, and zero otherwise. Panel (a) reports estimates for blame-only days, and Panel (b) reports estimates for credit-only days. The treatment indicator equals one for respondents interviewed in the seven days after a blame-only or credit-only day. Columns (1) and (2) use all respondents. Columns (3) and (4) restrict the sample to respondents who do not report having seen or heard political news on social media in the past week. Columns (5) and (6) restrict the sample to respondents who do report having seen or heard political news on social media in the past week. Columns (1), (3), and (5) absorb fixed effects for raw tweet volume on the rhetoric-only day, while columns (2), (4), and (6) control for the log of one plus tweet volume across the full comparison window. All specifications include district-by-month, date, partisan alignment, and respondent covariate fixed effects. Standard errors in parentheses are clustered at the congressional-district level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A15: Voting Intentions: 99th-Percentile Rhetoric Days

	(1)	(2)	(3)
	All Respondents	Split by Social Media Use	
		No Social Media Use	Social Media Use
(a) Blame			
Post-Blame	0.009* (0.005)	0.007 (0.010)	0.012** (0.006)
District x Month FE	✓	✓	✓
Date FE	✓	✓	✓
Partisan Alignment FE	✓	✓	✓
Respondent Covariate FE	✓	✓	✓
Observations	21,959	6,297	15,583
Clusters	315	309	315
(b) Credit			
Post-Credit	0.005 (0.004)	-0.000 (0.009)	0.005 (0.005)
District x Month FE	✓	✓	✓
Date FE	✓	✓	✓
Partisan Alignment FE	✓	✓	✓
Respondent Covariate FE	✓	✓	✓
Observations	29,921	8,496	21,330
Clusters	343	340	343

Notes: The table presents estimates of Equation 2 using an alternative definition of rhetoric days based on 99th-percentile blame and credit spike indicators. The outcome is an indicator equal to one if the respondent reports intending to vote for the politician, and zero otherwise. Panel (a) reports estimates for blame-spike days, and Panel (b) reports estimates for credit-spike days. The treatment indicator equals one for respondents interviewed in the seven days after a 99th-percentile blame or credit spike day. Column (1) uses all respondents. Column (2) restricts the sample to respondents who do not report having seen or heard political news on social media in the past week. Column (3) restricts the sample to respondents who do report having seen or heard political news on social media in the past week. All specifications include district-by-month, date, partisan alignment, and respondent covariate fixed effects. Standard errors in parentheses are clustered at the congressional-district level. *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

B Platform Changes

A natural concern is that the documented rise in causal rhetoric reflects changes in Twitter’s platform design rather than a change in how politicians communicate. Two platform changes are especially relevant during our sample period. On February 10, 2016, Twitter introduced algorithmic feed curation, replacing the strictly chronological ordering of tweets. On November 7, 2017, Twitter expanded the character limit from 140 to 280 characters. We find no evidence that either change generated a discontinuous increase in blame or credit tweets.

We first examine the introduction of algorithmic feed curation. The time-series evidence in Figure 3 already suggests that blame and credit do not increase noticeably in 2016. We test this more directly using a regression discontinuity design centered on February 10, 2016. The outcome is the daily share of tweets classified as blame or credit, measured in a symmetric 90-day window around the platform change. As shown in Figure B1, there is no evidence of a discontinuity at the cutoff. The estimated jump is -0.035 and statistically insignificant ($p = 0.529$). This result suggests that the introduction of algorithmic feed curation did not contribute meaningfully to the subsequent rise in causal rhetoric.

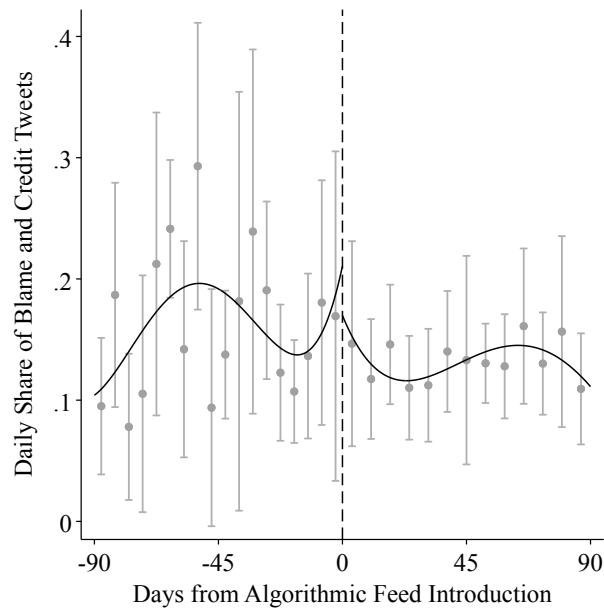
We next examine the expansion of the character limit. The relevant concern is that the original 140-character constraint may have disproportionately restricted blame or credit tweets, so that the later expansion mechanically enabled more causal rhetoric. Figure B2 plots the distribution of tweet length before November 7, 2017. The figure shows visible bunching in the upper part of the distribution, especially between 100 and 120 characters, indicating that many tweets were close to exhausting the original limit. Crucially, however, the pattern is similar across rhetorical categories. Blame, credit, and none tweets display comparable length distributions. If blame or credit had been especially constrained by the 140-character cap, one would expect stronger bunching near the limit for those categories. Instead, their distributions closely resemble that of none tweets.

We then test for a discontinuity around the character-limit expansion. Using the same regression discontinuity design centered on November 7, 2017, with the daily share of blame or credit tweets as the outcome and a symmetric 90-day window around the reform, Figure B3 shows no evidence of a jump at the cutoff. The estimated discontinuity is 0.056 and statistically insignificant ($p = 0.532$). The rise in causal rhetoric is therefore unlikely to be explained by the move from 140 to 280 characters.

Taken together, the evidence does not support the view that platform design mechanically generated the rise in causal rhetoric. Neither algorithmic feed curation nor the character-limit expansion coincides with a discontinuous increase in blame

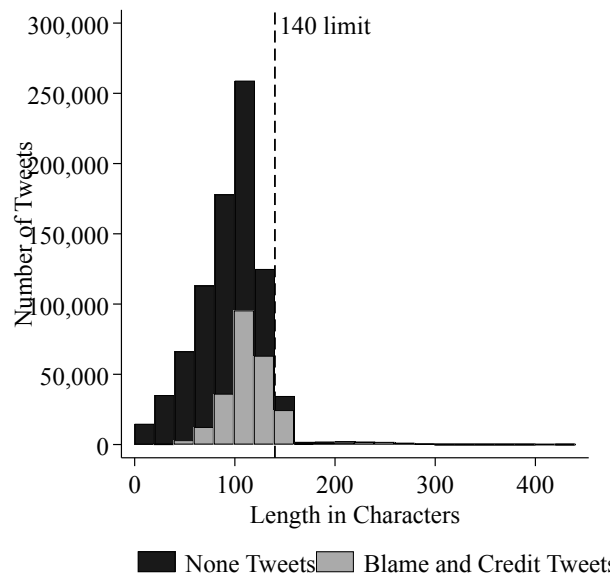
or credit tweets.

Figure B1: Regression Discontinuity Plot around Algorithmic Feed Introduction



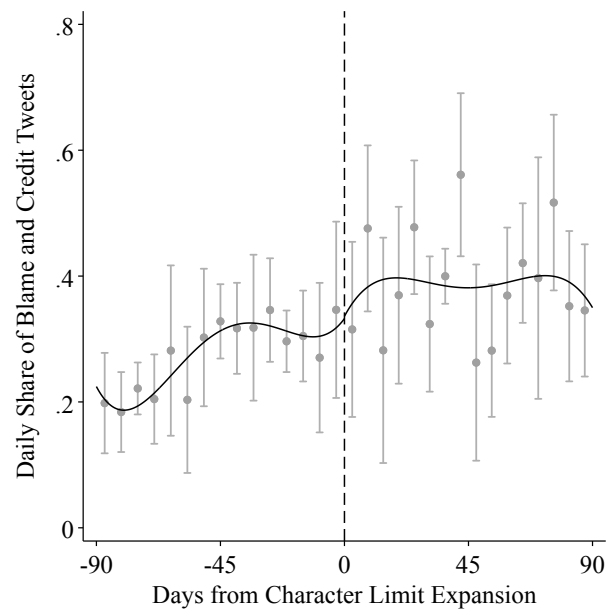
Notes: The figure reports the regression discontinuity plot around February 10, 2016, the date Twitter introduced algorithmic feed curation. The outcome is the daily share of tweets classified as blame or credit. The plot is produced using the `rdrobust` package of Calonico et al. (2017).

Figure B2: Distribution of Tweet Length before Character Limit Expansion



Notes: The figure shows the distribution of tweet length, measured in characters, for tweets posted before November 7, 2017, when Twitter expanded the character limit from 140 to 280 characters. The distribution is shown separately by rhetorical category.

Figure B3: Regression Discontinuity Plot around Character Limit Expansion



Notes: The figure reports the regression discontinuity plot around November 7, 2017, the date Twitter expanded the character limit from 140 to 280 characters. The outcome is the daily share of tweets classified as blame or credit. The plot is produced using the `rdrobust` package of [Calonico et al. \(2017\)](#).

C Trump and Causal Rhetoric

A natural concern is that the post-2016 rise in causal rhetoric reflects a Trump-style shift in congressional communication. The timing makes this concern natural: the sharp increase in blame begins around Trump's entry into office. If members of Congress absorbed Trump's rhetorical style, the rise in blame and credit could be a by-product of his influence on political language rather than a broader shift in congressional communication.

To assess this possibility, we collect Trump's tweets from the [Trump Twitter Archive](#) and classify them using the same blame–credit classifier employed throughout the paper. Because Trump's account was suspended in early 2021, these data end in 2020. We therefore restrict the analysis in this appendix to the period through 2020. This restriction is not central for our purpose, since causal rhetoric had already become substantially more common in congressional communication by then. We compare Trump's rhetorical style with that of members of Congress over the same period.

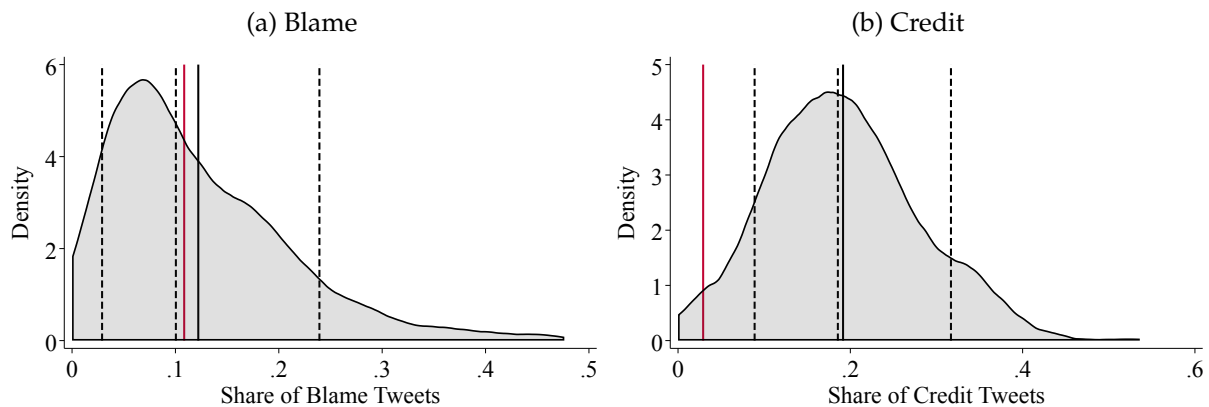
Trump is not an outlier in the use of causal rhetoric. For each member of Congress, we compute the share of tweets classified as blame and credit over the sample period and compare the resulting distributions with Trump's corresponding shares. As shown in Figures [C1a](#) and [C1b](#), Trump's use of blame is close to the congressional average—slightly above the median but slightly below the mean—while his use of credit is unusually low, falling below the 10th percentile of the congressional distribution. In this sense, Trump's rhetoric is not characterized by unusually intensive use of either blame or credit.

Trump also does not appear to anticipate the rise in causal rhetoric over time. Figures [C2a](#) and [C2b](#) compare the yearly shares of blame and credit tweets for Trump and for members of Congress. For blame, Trump's use evolves similarly to the congressional series and exceeds it only in 2018, by about 5 percentage points. For credit, Trump's use remains consistently below the congressional average throughout. This pattern makes it difficult to attribute the broader rise in congressional causal rhetoric to the diffusion of Trump's own rhetorical style.

What distinguishes Trump's rhetoric is not causal communication, but affective messaging. We repeat the same distributional exercise for non-causal tweets with negative and positive sentiment. Figures [C3a](#) and [C3b](#) show that Trump is an outlier in the share of non-causal tweets with negative sentiment, lying above the 90th percentile of the congressional distribution. His share of non-causal tweets with positive sentiment is also elevated—above the median and close to the 90th percentile—but less extreme. Thus, Trump is distinctive primarily for relying heavily on affective messaging, especially non-causal negative rhetoric.

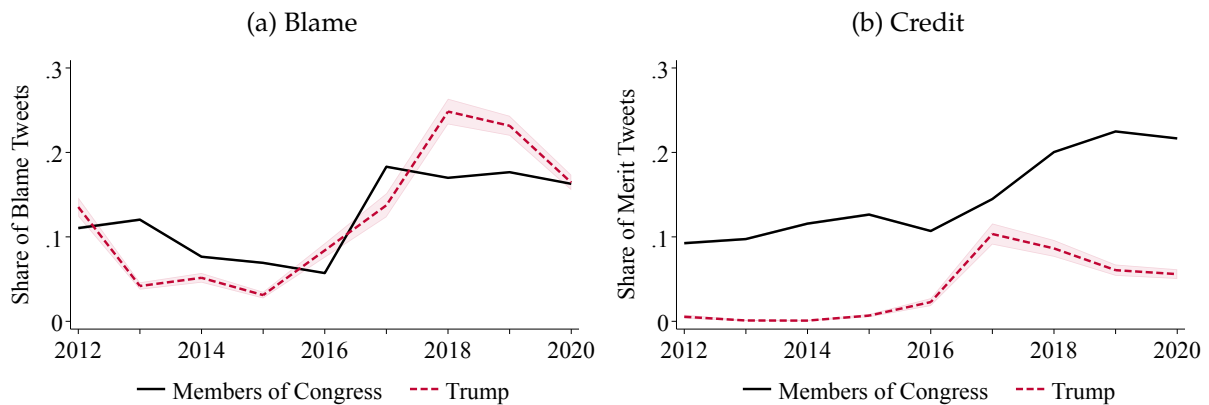
Overall, the evidence does not support the interpretation that members of Congress simply absorbed a Trump-style form of causal rhetoric after 2016. Trump is not unusually intensive in blame or credit. If anything, his distinctive rhetorical feature is a comparatively greater reliance on non-causal affective messaging, especially negative sentiment.

Figure C1: Trump and Causal Rhetoric



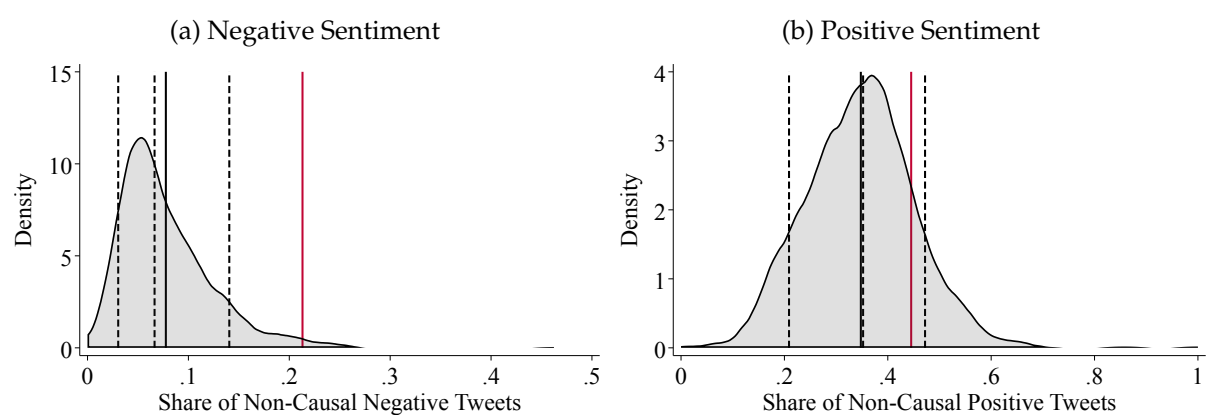
Notes: Each panel shows the distribution, across members of Congress, of the share of tweets classified as blame or credit over the 2012–2020 period. The solid black line marks the congressional mean, the dashed black lines mark the 10th, 50th, and 90th percentiles, and the solid red line marks Trump's corresponding share.

Figure C2: Trump and Causal Rhetoric over Time



Notes: Each panel shows the yearly share of tweets classified as blame or credit for members of Congress and for Trump over the 2012–2020 period. The solid black line reports the congressional average and the dashed red line reports Trump's corresponding share. Shaded areas represent 95 percent confidence intervals.

Figure C3: Trump and Non-Causal Affective Messaging



Notes: Each panel shows the distribution, across members of Congress, of the share of tweets that are non-causal and negative or positive in sentiment over the 2012–2020 period. The solid black line marks the congressional mean, the dashed black lines mark the 10th, 50th, and 90th percentiles, and the solid red line marks Trump's corresponding share.

D Persuasion Rate

To interpret the magnitude of our estimates, we calculate persuasion rates following [DellaVigna and Gentzkow \(2010\)](#):

$$f = \frac{y_t - y_c}{r \times (e_t - e_c)} \times \frac{1}{1 - y_0}. \quad (4)$$

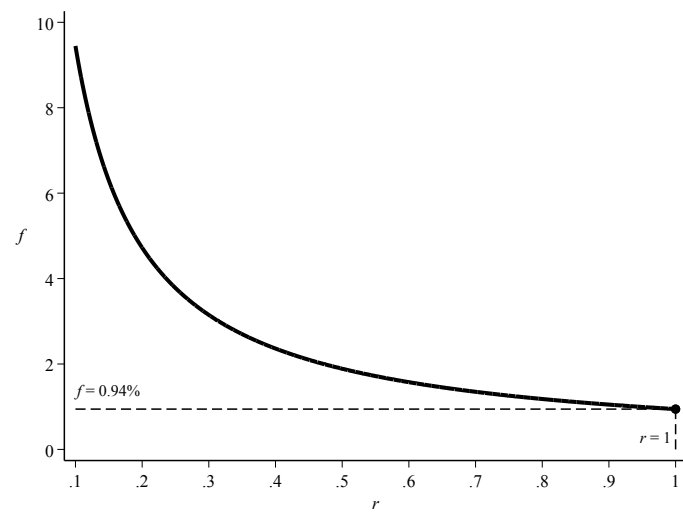
In this expression, $y_t - y_c$ is the difference in the share of the population donating between treated and untreated groups, y_0 is the share that would donate absent treatment, $e_t - e_c$ is the difference in exposure between treated and untreated groups, and r is a reach parameter.

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 the number of donors with respect to a specific rhetorical style and $\overline{\text{Users}}$ is average county log Twitter adoption, equal to 5.29 in our data. Since the estimated β for credit is close to zero and statistically insignificant, we focus on blame. We estimate $\beta = 0.003$. Following the literature, we assume $y_0 = y_c$, and we set $y_c = 0.16$ following [Boken et al. \(2023\)](#).

The term $r \times (e_t - e_c)$ requires additional discussion. While e_c is mechanically equal to zero, e_t represents the share of the average county that is on Twitter, equal to 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. Because [Boken et al. \(2023\)](#) focus on virality, they assume that a viral tweet is seen by everyone on Twitter, effectively setting $r = 1$. This is the same assumption that leads to the 0.94% persuasion rate reported in the main text, which should be interpreted as a lower bound.

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 that they see all of them, would imply $r = 0.4$ and therefore $f = 2.36\%$. To be transparent about this sensitivity, Figure D1 reports how the implied persuasion rate changes as the assumed reach parameter r varies. The implied magnitude remains meaningfully interpretable as r decreases: even at $r = 0.1$, the implied persuasion rate is $f \approx 9.45\%$, smaller than the Fox News effect documented in [DellaVigna and Kaplan \(2007\)](#).

Figure D1: Blame's Persuasion Rate and Reach



Notes: The figure plots the implied persuasion rate f for different values of the reach parameter $r \in [0.1, 1]$, holding fixed the remaining inputs used in the calibration described in this appendix.

E Blame and Amplification

This appendix provides additional details on the relationship between rhetorical style and amplification discussed in Section 5. We proceed in two steps. First, we describe the availability of engagement data and compare the corresponding subsample to the full tweet corpus. Second, we show that blame tweets receive more retweets and are disproportionately likely to appear in the upper tail of the retweet distribution, whereas credit tweets show no comparable pattern.

Engagement data are available only for a subsample of the tweets used in the main analysis. Retweet counts are observed through March 2020 and cover approximately half of all tweets in the sample over that period. Table E1 compares the full dataset with the engagement subsample. At the politician level, tweets with engagement data come from politicians who are, on average, about one year older and 2 percentage points more likely to be Republican. At the tweet level, differences are likewise modest: tweets with engagement data exhibit a slightly higher share of credit tweets, by about 4 percentage points, while the remaining dimensions are closely aligned. Overall, these comparisons suggest that the engagement subsample is broadly representative of the underlying sample, reducing concerns that the results below are driven by selection.

To study how blame and credit relate to engagement, we estimate the following specification:

$$r_{ipdt} = \alpha + \beta_1 \text{Blame}_i + \beta_2 \text{Credit}_i + \lambda_{p \times d} + \eta_{t \times d} + \varepsilon_{ipdt}, \quad (5)$$

where r_{ipdt} is the standardized number of retweets received by tweet i , posted by politician p on date d and topic t ; Blame_i and Credit_i are indicators equal to one if tweet i is classified as blame or credit, respectively; $\lambda_{p \times d}$ denotes politician-by-date fixed effects; and $\eta_{t \times d}$ denotes topic-by-date fixed effects. Standard errors are clustered at the politician level. The identifying comparison is therefore within politician-date, comparing tweets posted by the same politician on the same day, while also absorbing date-specific differences across topics. Figure E1 reports the estimates. Blame tweets receive, on average, about 0.15 standard deviations more retweets than tweets not classified as blame or credit. By contrast, credit tweets do not receive significantly more retweets than tweets not classified as blame or credit.

We next ask whether this relationship varies across the distribution of retweets. To do so, we re-estimate Equation 5 ten times, each time using as the dependent variable an indicator for whether the tweet falls into a given decile of the retweet distribution. Each regression therefore captures how blame and credit are associated with the prob-

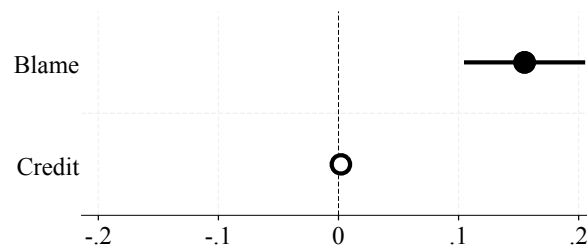
ability of appearing in different parts of the retweet distribution. Figure E2 shows that blame tweets are significantly less likely to appear in the lower deciles and increasingly more likely to appear in the right tail, especially in the top decile. Credit tweets, by contrast, are somewhat more likely to appear just above the median, but show no comparable tendency to reach the upper tail of the distribution.

Table E1: 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 Credit 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	2,322,957	1,178,861		

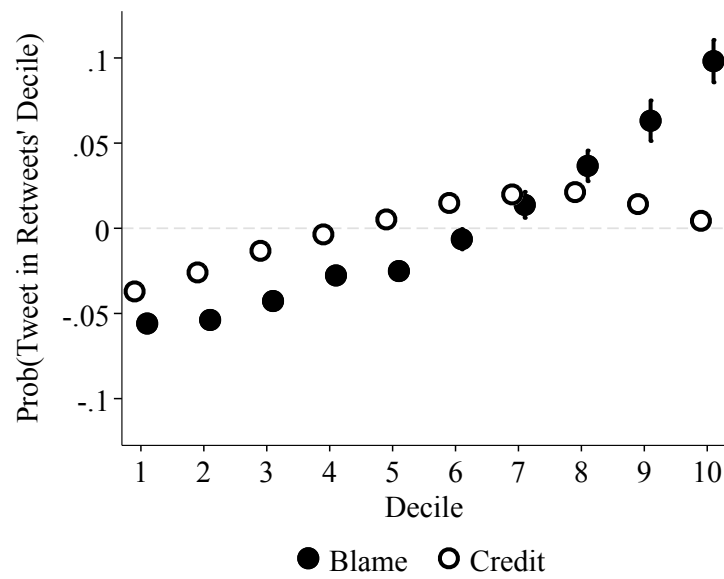
Notes: Means and standard errors in parentheses. The last column reports the p-value from a two-sided t-test of equality of means.

Figure E1: Blame, Credit, and Retweets



Notes: The figure presents the estimates of Equation 5. The outcome is the standardized number of retweets. Bars represent 95 percent confidence intervals computed with standard errors clustered at the politician level.

Figure E2: Blame, Credit, and Retweets' Distribution



Notes: The figure reports estimates from regressions that use, as the dependent variable, an indicator for whether a tweet falls into each decile of the retweet distribution. Bars represent 95 percent confidence intervals computed with standard errors clustered at the politician level.

F Assessing Economic Causal Rhetoric

This Appendix provides additional details on the construction and classification of the economic-tweet sample used in Section 6. It describes the sample, the coding protocol, and the implementation of the AI-assisted classification procedure, and reports illustrative examples.

The sample used in Section 6 consists of tweets classified as blame or credit that our topic classification identifies as being about the economy. This is the same topic-classification procedure used throughout the paper and discussed in Section 4. Applying that procedure to tweets classified as blame or credit yields 413,836 economic tweets. We do not impose any additional exclusions before classification. All such tweets are submitted to the AI-assisted coding procedure described below. A small number of requests fail because the model does not return a complete output in the required format. In practice, this concerns 379 tweets, corresponding to about 0.09 percent of the economic-tweet sample. These observations are therefore excluded from the analysis.

Each tweet is classified according to the dominant causal claim it conveys, if any. The coding task proceeds in three steps. First, for each tweet we extract the main causal claim. If no causal claim is present, the tweet is coded as *na*. Second, we identify the implied economic mechanism underlying that claim. Third, we classify the claim according to the extent to which it is supported by mainstream macroeconomic reasoning. The classifier assigns one of five raw labels: *standard*, *minor qualified*, *major qualified*, *unsupported*, and *na*. Observations coded as *na* are excluded from the analysis.²⁴

We implement the classification procedure using GPT-5 mini in batch mode under a fixed prompt and output schema. We instruct the model to rely on standard mainstream macroeconomic reasoning at the core undergraduate level because the objective of the exercise is not to adjudicate frontier debates in macroeconomics, but to assess whether a claim is supported by the settled core of economic knowledge. Using this benchmark keeps the coding transparent and reproducible, and avoids making classifications hinge on areas where expert disagreement remains substantial. The model is instructed to act as a careful research coder, to follow the codebook exactly, and to evaluate claims using standard mainstream macroeconomic reasoning. The system prompt reads as follows:

²⁴The *na* category accounts for about 14% of the economic-tweet sample. This is in line with the performance of our classifier. In the validation set, the classifier attains an average precision of about 0.84, implying that roughly one sixth of tweets classified as blame or credit should not in fact belong to those categories. Against this benchmark, the 14% share of tweets reassigned to *na* under the stricter macroeconomic coding procedure appears broadly consistent.

You are a careful research coder.
Follow the XML codebook exactly.
Use standard mainstream bachelor-level macroeconomic reasoning.
<codebook>

The full codebook is separately reported in Listing F1. Then, the user prompt asks the model to classify a tweet by passing the tweet identifier, text, and date in structured form.

Table F1 reports the distribution of labels in the economic-tweet sample, both pooled and separately for blame and credit tweets. Table F2 reports illustrative examples of tweets assigned to each classification category.

Table F1: Distribution of Labels

	Pooled	Split	
		Blame	Credit
Supported	9.23	8.54	9.79
Minor Qualified	61.76	46.07	74.54
Major Qualified	28.12	43.62	15.50
Unsupported	0.89	1.77	0.17

Notes: The table reports the distribution of macroeconomic-support labels in the economic-tweet sample. The pooled column reports the shares across all tweets that our topic classification identifies as being about the economy and that receive a non-*na* label under the macroeconomic coding procedure. The split columns report the same shares separately for tweets classified as blame and credit.

Table F2: Sample Tweets

Tweet	Label	Justification
I voted for additional direct payments in May and October, and tomorrow Ill vote again for \$2,000 checks to help those in need. \$600 is not enough, and Ill keep fighting for more. But thats no reason to stop a relief bill that will help people who are struggling. #SignTheBill	Standard	Cash transfers increase disposable income and consumption, easing financial hardship; standard macro effect under ordinary assumptions without major qualifiers.
Congress should pass my bill to get rid of the federal gas tax for the rest of the year. This is one step we can take right now to bring down high gas prices.	Minor Qualified	Recognized supply-side price effect; pass-through to consumer prices is standard but magnitude and timing depend on oil market, state taxes, retailer margins, and demand; temporary cut may have limited impact.
Biden's \$6T budget is reckless spending that doubles down on his far-left policies. Americans can not move forward if Biden's big government lowers paychecks, raises taxes \$ harms our economy. Republicans have a way forward. URL	Major Qualified	Claim links big fiscal package to lower pay and harm; mainstream macro sees fiscal policy can affect income, taxes, labor incentives and crowding out, but effects depend on composition, financing, monetary offset, labor market frictions and timing.
Extending the current spending levels is better than a government shutdown, but it is still bad policy because it damages our national security and wastes taxpayer dollars.	Unsupported	Tweet links extension of current spending to damage to national security and waste, but mainstream macro sees neutral/ambiguous macro effect of continuing spending; claims rely on program-specific allocation and micro/defense budget details.

Notes: The table reports illustrative examples of economic tweets assigned to each macroeconomic-support category under the AI-assisted coding procedure described in this Appendix. For each tweet, we report the assigned label together with the corresponding justification.

Listing F1: Codebook

```
<codebook name="macro_tweet_classification">
<task>
```

Classify the dominant economic causal claim in one political tweet according to how far mainstream bachelor-level macroeconomics would support it under the tweet's baseline interpretation: fully under ordinary assumptions, with only ordinary qualifications, with substantial but still mainstream qualifications, not plausibly under ordinary mainstream conditions, or not at all because no meaningful economic causal claim is present.

</task>

<input>

<item name="tweet_id">Unique identifier of the tweet.</item>

<item name="tweet_text">Text of the tweet to classify.</item>

<item name="tweet_date">Date of the tweet.</item>

</input>

<claim_rules>

<rule>Code the dominant economic causal claim in the tweet.</rule>

<rule>Code causal logic, not rhetoric, morality, persuasion, or political desirability.</rule>

<rule>Interpret rhetorical language as its most reasonable economic claim, but do not invent missing detail.</rule>

<definition>A meaningful economic causal claim states or clearly implies that a policy, political, or economic factor affects an economic outcome.</definition>

<support_scope_rules>

<rule>Do not assign "na" because the claim is difficult to verify empirically, because the tweet refers to a specific policy or episode, or because the named policy is not discussed in textbooks.</rule>

<rule>If no meaningful economic causal claim is present, assign "na".</rule>

<rule>If a meaningful economic causal claim is present, assign one of "standard", "minor_qualified", "major_qualified", or "unsupported".</rule>

</support_scope_rules>

</claim_rules>

<interpretation_rules>

<definition name="baseline_interpretation">

The most reasonable ordinary reading of the tweet's dominant causal claim, without strengthening or weakening it through charitable reinterpretation.

</definition>

<definition name="ordinary_economic_assumptions">

Familiar background assumptions that normally accompany causal reasoning in mainstream undergraduate macroeconomics.

</definition>

<definition name="implied_economic_mechanism">

The abstract economic channel connecting the cause and the outcome, stated independently of the tweet's policy-specific, partisan, geographic, or branded language.

</definition>

<rule>Translate the causal claim, if any, into the implied economic mechanism before evaluating support.</rule>

<rule>Evaluate the tweet at the level of the economic mechanism, not at the level of the named policy label.</rule>

<rule>Use the specific policy name only to identify the implied economic mechanism, not as the object of support.</rule>

<rule>Do not invent a mechanism that introduces a new causal bridge not reasonably expressed in the tweet.</rule>

```

<rule>Do not rescue a claim merely because some logically possible economic story can be
    invented for it.</rule>
</interpretation_rules>

<knowledge_rules>
<rule>Use standard mainstream bachelor-level macroeconomics as taught in introductory and
    intermediate undergraduate courses.</rule>
<rule>Do not rely on niche, heterodox, or highly specialized theories unless they are part
    of standard undergraduate macroeconomic instruction.</rule>
<rule>Evaluate the tweet's causal logic at the level of the implied economic mechanism, not
    at the level of the named policy, politician, bill, slogan, or episode.</rule>
<rule>Do not use hindsight, ex post outcomes, or later historical developments to evaluate
    the claim.</rule>
<rule>Do not fact-check whether the named policy or event actually produced the stated
    outcome; assess whether mainstream macroeconomics would support the causal logic.</rule>
</knowledge_rules>

<temporal_context>
<rule>The tweet date may be used only to identify contemporaneous context and the likely
    target of the claim.</rule>
<constraint>Do not use later events or ex post outcomes to evaluate the claim.</constraint>
</temporal_context>

<labels>
<label name="standard">
<definition>
    Use "standard" when mainstream bachelor-level macroeconomics would support the tweet's
    dominant causal claim under ordinary economic assumptions. The claim may still be
    simplified, but no relevant additional qualification is needed.
</definition>
<inclusion_criteria>
<item>The implied mechanism is recognized in mainstream bachelor-level macroeconomics.</
    item>
<item>The claim is supportable under ordinary, familiar assumptions in undergraduate
    macroeconomics.</item>
<item>No relevant contextual qualification or scope restriction is needed.</item>
</inclusion_criteria>
<exclusion_criteria>
<item>The claim is supportable only if important contextual conditions or scope
    restrictions do substantial work.</item>
<item>No plausible mainstream macroeconomic condition would support the claim.</item>
</exclusion_criteria>
</label>

<label name="minor_qualified">
<definition>
    Use "minor_qualified" when the tweet's dominant causal claim maps onto a recognized
    mainstream macroeconomic mechanism and is broadly in line with its standard direction,
    but support depends on familiar contextual qualifications, scope restrictions, or
    ordinary assumptions.
</definition>
<inclusion_criteria>
<item>The implied mechanism is recognized in mainstream undergraduate macroeconomics.</
    item>
<item>The claim's direction is broadly consistent with the standard prediction.</item>
<item>Support requires ordinary caveats about magnitude, timing, pass-through, openness,
    slack, or policy context.</item>

```

```

    <item>The qualifications refine the claim but do not substantially rewrite its baseline
        meaning.</item>
</inclusion_criteria>
<exclusion_criteria>
    <item>The claim is supportable under ordinary assumptions without meaningful qualification
        .</item>
    <item>Support requires major unstated offsets or a reversal of the standard directional
        effect.</item>
    <item>No ordinary plausible mainstream condition would support the claim.</item>
</exclusion_criteria>
</label>

<label name="major_qualified">
<definition>
    Use "major_qualified" when the tweet's dominant causal claim can be supported by a
        recognized mainstream macroeconomic mechanism only if substantial unstated conditions,
        limiting assumptions, or offsetting forces hold, but these are still not so narrow or
        contrarian as to make the claim unsupported.
</definition>
<inclusion_criteria>
    <item>The implied mechanism is still recognizable within mainstream undergraduate
        macroeconomics.</item>
    <item>Support requires substantial qualifications that do heavy interpretive work.</item>
    <item>The baseline reading is not the standard prediction, but a mainstream route to
        support still exists under non-remote conditions.</item>
    <item>The claim may require specific policy composition, monetary offset, special
        incidence, or other important conditions not stated in the tweet.</item>
</inclusion_criteria>
<exclusion_criteria>
    <item>The claim is supportable under ordinary assumptions or only familiar caveats.</item>
    <item>Support would require narrow, unusual, strongly contrarian, or specially engineered
        assumptions.</item>
    <item>The model would need to invent a new causal bridge not reasonably expressed in the
        tweet.</item>
</exclusion_criteria>
</label>

<label name="unsupported">
<definition>
    Use "unsupported" when no ordinary plausible mainstream undergraduate macroeconomic route
        supports the tweet's dominant causal claim under its baseline interpretation, or when
        support would require narrow, unusual, strongly contrarian, or specially engineered
        assumptions.
</definition>
<inclusion_criteria>
    <item>No ordinary plausible mainstream macroeconomic condition would support the claimed
        relationship.</item>
    <item>The claim's direction conflicts with the standard prediction and can be defended
        only by unusual or remote offsets.</item>
    <item>The implied mechanism is not recognized in mainstream undergraduate macroeconomics
        .</item>
    <item>Any defense would require rescuing the tweet beyond its ordinary meaning.</item>
</inclusion_criteria>
</label>

<label name="na">
<definition>

```

```

    Use "na" when the tweet contains no meaningful economic causal claim.
</definition>
<inclusion_criteria>
    <item>The tweet is descriptive, moral, partisan, rhetorical, or evaluative without a
        meaningful economic causal claim.</item>
    <item>No dominant economic causal claim can be extracted from the tweet.</item>
    <item>Any implied causal reading would require inventing details not reasonably expressed
        in the tweet.</item>
</inclusion_criteria>
<exclusion_criteria>
    <item>A recognizable economic causal claim is present.</item>
    <item>The tweet can be classified as "standard", "minor_qualified", "major_qualified", or
        "unsupported".</item>
</exclusion_criteria>
</label>
</labels>

<classification_principles>
    <rule>The labels do not measure empirical truth; they measure how far mainstream
        undergraduate macroeconomics would go in supporting the dominant causal claim under its
        baseline interpretation.</rule>
    <rule>Do not reward a claim merely because some theoretical possibility exists.</rule>
    <rule>Use "standard" when support follows under ordinary assumptions without meaningful
        qualification.</rule>
    <rule>Use "minor_qualified" when support requires familiar and ordinary qualifications that
        refine, but do not substantially alter, the tweet's baseline meaning.</rule>
    <rule>Use "major_qualified" when support requires substantial unstated qualifications or
        offsetting conditions that still remain within mainstream undergraduate macroeconomics
        and are not remote or contrarian.</rule>
    <rule>Use "unsupported" when no ordinary plausible mainstream macroeconomic condition would
        support the claim, or when support would require narrow, unusual, strongly contrarian,
        or specially engineered assumptions.</rule>
    <rule>Use "na" when no meaningful economic causal claim is present.</rule>
</classification_principles>

<language_rules>
    <rule>Write all output fields in English only.</rule>
    <rule>Do not use any non-English characters.</rule>
</language_rules>

<coding_procedure>
    <step number="1">Determine whether the tweet contains a meaningful economic causal claim. If
        not, assign "na", leave all non-applicable fields empty, and stop coding.</step>
    <step number="2">Extract the dominant causal claim in one short sentence.</step>
    <step number="3">Identify the implied economic mechanism.</step>
    <step number="4">Ask whether mainstream bachelor-level macroeconomics supports the claim
        under ordinary assumptions, supports it only with important qualifications, or does not
        plausibly support it.</step>
    <step number="5">Assign exactly one classification label: "standard", "qualified", "
        unsupported", or "na".</step>
    <step number="6">Provide a brief justification grounded in mainstream macroeconomic
        reasoning, emphasizing why support is standard, minor_qualified, major_qualified, or
        unsupported.</step>
</coding_procedure>

<output>
    <format>Return exactly one valid JSON object.</format>

```

```

<requirements>
  <item>Return only the JSON object.</item>
  <item>Use empty strings where a field is not applicable.</item>
</requirements>
<fields>
  <field name="tweet_id">Unique identifier of the tweet.</field>
  <field name="causal_claim_extracted">The dominant causal claim in one short sentence. If
    none, return an empty string. (max 100 characters)</field>
  <field name="mechanism_identified">The implied economic mechanism in one short sentence. If
    none, return an empty string. (max 100 characters)</field>
  <field name="classification">One of: standard, minor_qualified, major_qualified,
    unsupported, na.</field>
  <field name="justification">A concise, schematic justification grounded in mainstream
    macroeconomic reasoning. Use short phrases. (max 250 characters)</field>
</fields>
</output>
</codebook>

```