Virality:
What Makes Narratives Go Viral, and Does it Matter?*

Kai Gehring† and Matteo Grigoletto‡

October 8, 2025

Abstract

The effectiveness of political narratives as a communication technology depends on their virality and on the persuasiveness of single narrative exposure. To analyze narratives empirically, we introduce the political narrative framework and a pipeline for its measurement using large language models (LLMs). The framework captures the essence of a narrative by its characters, who are either neutral or cast in one of three drama triangle roles: hero, villain, or victim. Using 1.15 million U.S. climate policy tweets from 2010–2021, we find that political narratives are consistently more viral than comparable neutral tweets. This result is robust to conditioning on a rich set of fixed effects, author characteristics, language metrics and emotionality. Hero roles increase virality by 56%, but the biggest virality boost stems from using villain roles (152%) and from combining other roles with villain characters. To examine the persuasiveness of single exposure to some of the most frequent and viral character-role combinations, we use three pre-registered online experiments with 3000 participants. The results show that narrative exposure influences beliefs and revealed preferences about a character, while a single exposure is not sufficient to move support for specific policies. Political narratives also lead to consistently higher memory of the narrative characters and their roles, while memory of objective facts is not improved. Taken together, the political narrative framework provides a measure that moves beyond emotions and linguistic features, helps to explain virality, and is linked to shifts in beliefs, revealed preferences, and memory.

Keywords: Narrative economics, climate change policy, virality, economics of social media, political economy, media economics, text-as-data, machine learning, large language models.

JEL Classification: C80, D72, H10, L82, P16, Q54, Z1

*We thank Gustav Pirich, Lina Goetze, and Lorenz Gschwent for excellent research assistance; all remaining mistakes are our own. We are grateful for helpful suggestions from Toke Aidt, Peter Andre, Elliott Ash, Bruno Caprettini, Milena Djourelova, Ruben Durante, Vera Eichenauer, Ruben Enikolopov, Ingar Haaland, Gaël Le Mens, Pierre-Guillaum Meon, Christopher Roth, Paul Schaudt, Marta Serra-Garcia, Johannes Wohlfahrt, Joachim Voth, as well as from participants and discussants at the Barcelona Summer Forum 2025, the Bolzano workshop on historical economics 2023, the Memories, Narratives and Beliefs Workshop in Riednau, the 4th Monash-Warwick-Zurich Text-as-Data Workshop, the EARE 2023 in Cyprus, the ifo Social Media in Economics conference (Venice), the 6th ifo Economics of Media Bias Workshop, the International Institute of Public Finance Conference, the European Public Choice Society Conference (EPCS) 2023 in Hannover, the EPCS 2025 in Riga, the Silvaplana Workshop of Political Economy, the NLP for Climate Politics Workshop, the 2023 Narrative Economics workshop in Zurich, and the Beyond Basic Questions (BBQ) Workshop in Stuttgart 2024. We also thank audiences and discussants at seminars and guest lectures at Uppsala University, at Friedrich-Alexander University Erlangen-Nuremberg,the University of Bolzano, ETH Zurich, Paris Dauphine, Hamburg University, and the University of Bern. All remaining mistakes are ours. A previous version of this paper was titled "Analyzing Climate Change Policy Narratives with the Character-Role Narrative Framework" and circulated as CESifo Working Paper No. 10429 in 2023. The study obtained ethics approval from the University of Bern Faculty of Business, Economics and Social Sciences ethics board. It was pre-registered with AsPredicted (see pre-registration for the first, second, and third experiment).

†University of Bern, Wyss Academy at the University of Bern, CESifo, e-mail: *mail@kai-gehring.net*

‡University of Bern, Wyss Academy at the University of Bern, e-mail: *matteo.grigoletto@unibe.ch*

1 Introduction

Politics, broadly conceived, is the social process of settling conflicts over collective decisions – "through words and persuasion, and not through force" (Hannah Arendt). Such conflicts occur every day in legislatures, city councils, board meetings, workplaces, friends' circles and families. Narratives can be regarded as the tailored communication technology that actors employ to succeed in that process. Political narratives compress complex information about human or instrument characters into the three archetypal roles of the drama triangle – hero, villain, victim – that can be processed, remembered and, crucially, shared easily. Hence, political narratives have an efficiency advantage in information markets when attention is limited and information processing costly. This article investigates (i.) which features enhance the virality of a political narrative and (ii.) the impact of political narratives on beliefs, preferences and memory.

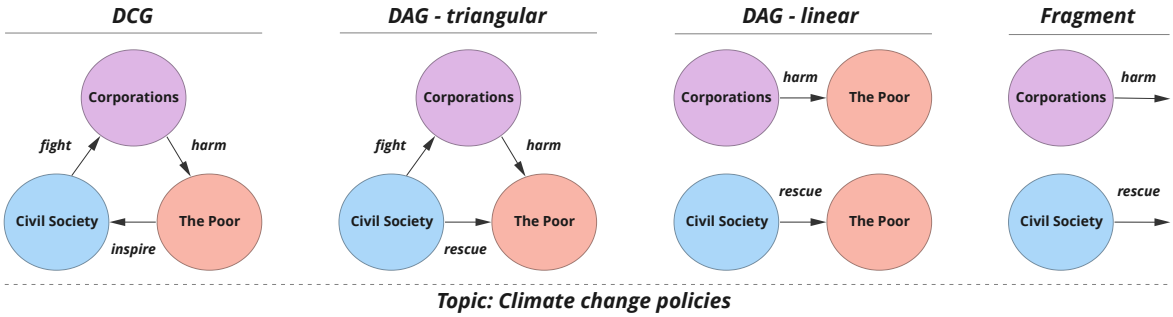
In popular science books (Harari 2014; Shiller 2020), but also increasingly in economics (e.g., Bénabou, Falk, and Tirole 2020; Bursztyn et al. 2023; Esposito et al. 2023) narratives are now generally recognized as a key communication technology that influences human preferences, beliefs, and decisions. Andre et al. (2025), Eliaz and Spiegler (2020), and Kendall and Charles (2022) started analyzing narratives systematically as causal sequences, and Barron and Fries (2024) evaluate the persuasive power of such causal narratives. However, the importance of narratives does not solely stem from their persuasive power, but as Shiller (2017) highlights, as well as from their ability to go viral. When studying virality in applications with real-world data, focusing on causal sequences has important limitations. In much of human communication – be it news, social media, or speeches – narratives are often fragmented and sometimes lack explicit causal structure, or rely on emotions, framing, tone, and dynamics between characters. To complement these existing studies, we conceptualize a broader class of political narratives, integrating insights from various other disciplines, as well as a measurement pipeline using large language models (LLMs).

This enables us to study the two key features of narratives: virality and persuasive power. We study virality where it is most visible: the retweet metrics on the social media platform Twitter/X. We analyze 1.15 million climate change policy tweets over the 2010 to 2021 period, encoding ten predefined human and instrument characters as being in a neutral or a hero-villain-victim role. Political narratives are markedly more viral than otherwise similar messages, even after controlling for author characteristics, text quality and emotions. Villain and hero roles produce the largest individual boost and combinations that add extra villain characters further raise virality. In complementary, pre-registered survey experiments with 3000 respondents, political narrative exposure shifts beliefs and incentivized donations in the predicted direction, yet leaves stated policy positions largely unchanged. The narrative treatments also improve recall of characters, while memory for numeric facts does not improve. Hence, the power of political narratives stems from both repeated exposure to viral narratives, as well as from individual persuasiveness and improved memory.

But let's begin by better illustrating what constitutes a political narrative and how to move from concept to empirical measurement. The purpose of a political narrative is influencing perceptions,

beliefs, and preferences about the characters contained in the narrative – hence the term "political". Political narratives exert their influence by depicting characters in one of three archetypal roles from the so-called "drama triangle" (Karpman 1968) – hero, villain, or victim. Characters may be human – individuals or collective actors such as corporations, states, or movements – or instrument, denoting policies, laws, or technologies. Consider as an example, the following quote by climate activist Greta Thunberg:

> *"**Global greenhouse emissions** are still on the rise, oil production is soaring and energy companies are making sky-high profits while countless people struggle to pay their bills. [...] A critical mass of people – especially younger people – are demanding change and will no longer tolerate the procrastination, denial and complacency that created this state of emergency."*
>
> Greta Thunberg, *The New Statesman*, 19^{th} Oct. 2022



For a given topic – here, climate-change policy (*global greenhouse emissions*) – the passage shows that naming the key characters and assigning them roles captures the essence of a political narrative. Corporations (*energy companies*), the poor (*countless people*), and civil society (*younger people, activists*) are the characters, depicted as the nodes in the graphs above. The dynamic (a)cyclical graphs (DCG/DAG) below the quote show that assigning causal arrows between characters may often be ambiguous in real texts, moreover, any causal representation presupposes that the relevant nodes have been defined and measured. By contrast, role assignment is typically clearer and can be coded directly: in this example, corporations are cast as villain, the poor as victim, and civil society as hero. This character-role coding recovers both grand narratives with multiple character-roles (left DCG/DAG graphs) and shorter fragments that are more common in natural text (right fragments of DCG/DAG graphs). The definition and measurement of political narratives, therefore, reduce to specifying the topic and characters, and coding for each character whether it appears as neutral or cast as hero, villain, or victim.

> *"A political narrative is identified by (i) its topic, (ii) its characters, and (iii) by having at least one character cast in a drama triangle role: hero, villain, or victim."*

We provide a general pipeline, implemented in Python, to measure political narratives with LLMs for any topic and user-defined set of characters. The pipeline prompts an LLM to (i) detect the topic, (ii) flag the presence of specified characters, and (iii) assign roles where present, returning

structured outputs that feed directly into statistical analysis. The ubiquity of the archetypal roles of the drama triangle in human texts ensures that even cheap and established LLMs – we use GPT-4o-mini from OpenAI – perform well even with simple one-shot prompting. Our pipeline returns a compact panel of variables – one row per tweet and columns indicating character presence and role assignment – that is straightforward to use for further analysis.

To analyze what makes political narratives go viral, we choose climate change policy as a topic and focus on social networks, specifically on the social network Twitter/X. Climate change policy is a suitable topic for this inquiry: its long time horizon denies participants quick feedback on outcomes, leaving narratives largely unverified in the short run and therefore dependent on their virality for influence. Because the underlying issue is characterized by deferred pay-offs and distributional conflict, political narratives have ample room to shift preferences and beliefs by assigning credit or blame without near-term falsification. Social media has been the focus of much of recent research in economics (see overview in Aridor et al. 2024), and has the advantage of offering clear metrics of virality. We select Twitter as a platform that became a central arena for agenda-setting and narrative contestation in Western politics (Halberstam and Knight 2016; Acemoglu, Hassan, and Tahoun 2018; Macaulay and Song 2022) over our sample period from 2010-2021. Twitter allows us to cover more than a decade of US discourse, tracking how political narratives evolve alongside the shifting public support patterns reported in cross-country surveys (Andre et al. 2024; Dechezleprêtre et al. 2025).

To explore our framework's capabilities, we pre-define ten characters, five human (e.g., corporations, US people) and five instrument (e.g., fossil industry, green technology, regulations) characters based on the literature, exploratory text analysis (e.g. topic models, word clouds) and our own reading of a large set of example tweets. Using climate-policy keywords following Oehl, Schaffer, and Bernauer (2017), we collect over three million English-language tweets via the Twitter API, sampling every Saturday plus one randomly chosen day per month from 2010 to 2021. We further use the metadata to select tweets originating from the US, ending with roughly 1.15 million tweets. Then we apply our pipeline, using the GPT-4o-mini model from OpenAI, to first further validate whether a tweet fits the topic. For each in-topic tweet, we detect whether each pre-defined character is present and, if so, whether it appears neutrally or in a drama triangle role: hero, villain, or victim.

We then begin by examining key text features that could be relevant for the virality and persuasive power of political narratives. We distinguish sentiment valence (positive–negative) and discrete emotions (e.g., joy, fear) from language metrics (e.g. length, readability). In our sample, narratives tend to employ more emotional language, with each role relying on different types of emotions. Using a standard emotion lexicon, we find that hero narratives are characterized by expressions of joy and surprise, victim narratives evoke fear, and villain narratives generally convey more anger and disgust. At the same time, we find no major differences on the language metrics. For instance, narratives are not generally easier to read or more densely written. Emotions and sentiment are one channel through which role assignment is signaled; roles can also be conveyed by explicit character labels, causal language or other cues.

We proceed with descriptive results on political narrative frequency over time, which show marked shifts between 2020 and 2021. The Green Tech–Hero character-role is most frequent, followed by Fossil Industry–Villain and Corporations–Villain, while the US public appears often as both hero and victim, with notable changes over time. Beyond individual character–roles, we explore the co-occurrence within a tweet of multiple roles, which reveals additional details about the political narratives over that time period. Taken together, these patterns provide an initial map of how political narratives – measured through the character–roles and their combinations – evolve in the climate-policy discourse.

In our main analysis, we exploit the fact that the retweet metric on Twitter constitutes a natural proxy for virality.[1] We begin by showing that the distribution of retweets is highly skewed: approximately 80% of tweets receive no retweets at all, with a rather small percentage of tweets receiving a large share of retweets. Using Poisson Pseudo-Maximum Likelihood regression models (PPML) suitable for count outcomes like retweets, we first assess the impact of simply featuring any political narrative, i.e., the tweet concerns the topic and includes at least one character cast as hero, villain, or victim, compared to featuring characters only neutrally. The presence of a political narrative has a positive and highly statistically significant effect on virality, which persists after controlling for author characteristics and character fixed effects. On top of that, we then control for sentiment, emotions and language metrics, evaluating whether this virality premium is solely or mostly driven by emotionality and text quality. However, the positive relationship with virality remains clearly positive and significant, highlighting that character-role combinations capture something more fundamental about the texts.

Next, we use the more detailed information from our pipeline about characters, roles and their combinations to investigate the determinants of political narrative virality. Our results indicate that both hero and villain narratives consistently boost virality compared to neutral featuring of the same characters, while victim narratives do not exhibit a significant effect. In models that include all roles, villain narratives emerge as the primary driver of engagement, often amplifying the impact of hero narratives. These findings are initially derived from tweets featuring a single character-role. However, when we examine tweets with multiple roles, a clear pattern emerges: increased narrative complexity generally reduces virality, particularly when a tweet features a mix of different roles such as hero and victim or all three roles together. The only exceptions occur when a hero is paired with a villain or a victim with a villain – configurations that are more viral than a hero alone. Overall, the reinforcement of villain narratives appears to be the most influential factor in driving engagement.

Our observational analysis has several limitations. First, Twitter/X was widely used in the US across partisan lines during our sample period, but users do not systematically represent the general US population. Second, the climate change policy discussion in the US context is not identical with that in e.g. the European Union, necessitating further studies of other countries and settings. Third, future studies should explore whether similar patterns can be found for other topics.

[1] We use "likes" and "replies" as another measure of narrative success. Although not equivalent to retweets and al ess direct proxy for virality, we observe very similar patterns. All results shown in the appendix.

We view climate policy and Twitter/X as a valuable first setting; the measurement framework and pipeline are general and can be applied to other topics and corpora. Third, our analysis of virality is correlational: despite extensive controls and fixed effects, selection and confounding cannot be entirely ruled out. Fourth, platform ranking and the Twitter algorithm can itself shape what is seen and shared. Controlling for sentiment, emotions, and language features somehow limits the issue, but algorithmic amplification remains a residual concern that cannot be avoided when studying social networks. Fifth, automated accounts (bots) may affect diffusion dynamics; we apply standard filters for robustness tests, but acknowledge that social bots remain an endemic feature of social media.

To complement the observational analysis of virality, we conduct a series of online survey experiments studying the persuasive power of single exposure to specific political narratives. Specifically, we compare the effects of a treatment panel of social media posts containing a political narrative with an active control panel of comparable length and complexity that contains the same characters in a neutral role. We find that narratives shape beliefs to some degree when two hero characters are present, but much more strongly with two villain characters. Political narratives also significantly shift revealed preferences about a character, using Green tech as a character and a pro-green technology donation as a real-stakes outcome. Single political narrative exposure does not significantly shift concrete policy preferences, but it is plausible that there are cumulative effects of repeated exposure ("mere exposure effect") to highly viral narratives(see review in Montoya et al. 2017). Finally, political narratives do enhance memory: participants recall posts containing narratives better. However, what they remember better are the characters (and their roles), not objective facts embedded in the narrative.

Contributions: We develop the political narrative framework as a novel approach for analyzing narratives – a narrative is defined by its topic and its characters, with at least one character cast as hero, villain, or victim. This representation complements the "narratives-as-causal-sequences" approach in economics (Eliaz and Spiegler 2020; Andre et al. 2025; Kendall and Charles 2022; Barron and Fries 2024), and also captures widely used non-sequential narrative statements, while remaining compatible with DAG/DCG causal representations. Drawing on successful applications – though mostly qualitative – in political science (Terry 1997; Jones and McBeth 2010; Jones 2014; Jiangli 2020), sociology (Polletta et al. 2011; Merry 2016; O'Brien 2018), communication studies (Anker 2005; Gomez-Zara, Boon, and Birnbaum 2018), and literary studies (Fog et al. 2010), we bring this role-based taxonomy to economics to formalize a concise and parsimonious representation of political narratives. The framework is topic-agnostic, making it portable across domains of interest to economists whenever a topic and a set of salient characters can be specified.

Second, we provide a Python-based pipeline for empirical measurement that takes as inputs a topic and a user-defined list of human and instrument characters and returns, for each text unit, character presence and role (hero, villain, victim, neutral). Using pre-tested prompts and batch-level API calls, the pipeline produces a structured panel – one row per document and columns for character presence and role assignment – for further statistical analysis. The design captures a

broader set of narratives that appear in real text without requiring an explicit sequence (c.f. Akerlof and Snower 2016), and it aligns with DAG/DCG representations (e.g. Andre et al. 2025): any causal narrative estimation presupposes defined nodes, and our character–role mapping provides those nodes even when causal direction is ambiguous. Whereas prior innovative software packages like RELATIO (Ash, Gauthier, and Widmer 2024) extract subject–verb–object sequences and defers dimension reduction and aggregation of entities into characters usually to later stages, our approach moves that dimension reduction upfront by defining characters ex ante. What we define as "characters" represents a broad set of human and instrument entities, allowing researchers to adapt the scheme readily to their setting. In contrast to prior, mostly manual, approaches to encode causal narratives, our pipeline allows measuring political narratives at scale in large text corpora with widely available LLMs.[2] We release python code and batch scripts together with the pre-tested prompts, instructions and simple consistency rules (e.g., mutually exclusive roles per character)[3].

Third, our empirical results on the virality of political narratives contribute to the growing literature in economics empirically analyzing the impact of specific narratives, which we can also view through the lens of our framework. Esposito et al. (2023) demonstrate that a specific, revisionist political narrative depicting African Americans as the villain character in the US Civil War shaped political preferences and behavior. Bursztyn et al. (2023) show how exposure to specific narratives about the COVID pandemic emphasizing either China or the US democratic party as the villain changes beliefs and even high-stakes health-related behavior. Our results on virality show for a broad set of political narratives with the topic climate change policy which character-role combinations tend to spread more and thus could influence larger audiences. Virality has not been studied empirically in economics; the few related studies in marketing and communication find, for instance, that arousal and emotion correlate with the virality of news (Berger and Milkman 2012)[4]. Our approach allows much more systematic insights compared to examining a set of text features or different emotions by showing which character types (human vs. instrument) and which roles and role combinations are linked to higher virality on social media.

Fourth, beyond virality, our pre-registered experiments complement recent causal evidence on the persuasiveness of narratives . Andre et al. (2025) study causal narratives – explicit action → consequence – and show that such sequences can causally shift beliefs. Barron and Fries (2024) identify the mechanics of narrative persuasion, emphasizing sense-making and fit to facts. In contrast, we test political narratives defined by character–role assignments and find that single-exposure treatments systematically move beliefs and a real-stakes choice about the named character, even without specifying an explicit causal chain. Thus our evidence is complementary: it brings role-based persuasion – closer to how political narratives appear in practice – into a controlled setting,

[2]Similar to other recent paper that also employ LLMs or deep-learning models in innovative ways (e.g. Ash et al. 2021; Lagakos, Michalopoulos, and Voth 2025; Voth and Yanagizawa-Drott 2025) to encode long stories and images, we demonstrate how LLMs allow us to move beyond simple metrics like emotions and study more structural features of narratives systematically.

[3]Contact the authors for an early version of the package and pipeline, which will be released after publication.

[4]Caesmann et al. (2021) study virality and persuasion, but focusing on the virality of propaganda transmitted not through media, but through public events.

while the causal-sequence studies pinpoint mechanisms at a finer level of detail.

Fifth, our experiments identify a memory channel through which single-exposure to political narratives matters: characters are better recalled if they are cast in a drama triangle role, whereas there is no effect on numerical facts embedded in the narratives. Compared to Graeber, Roth, and Zimmermann (2024), who compare stories to statistics and manipulate cues directly in a very controlled setting, our experiments are closer to real political content as in Barrera et al. (2020), varying role assignment while holding format and objective information constant. Decomposing the effect of different roles shows how hero role assignment increases recall; adding villain roles increases recall of these characters even more strongly while crowding out the memory of hero characters. Consistent with the emerging literature on attention economics (Serra-Garcia 2025; Loewenstein and Wojtowicz 2025), scarce attention resources in settings with information overload means content needs to secure attention to be remembered. The drama triangle roles receive attention by activating existing heuristics and mental models of the world, and villain characters who pose a potential threat receive more attention than hero characters. These findings invite future research to test in more detail aspects like cue similarity and interference in political contexts, and an investigation of how senders strategically use role assignments and role combinations to shape what is remembered.

Sixth, our results complement the growing, often survey-based, literature on the political economy of climate change. Andre et al. (2024) show that correcting misperceived norms increases support for climate change policies, and Dechezleprêtre et al. (2025) provide cross-country, experimental evidence that perceptions and ideology are more important than knowledge and understanding in shaping climate policy preferences. Djourelova et al. (2024) show that experiences of natural disasters and their media coverage affect climate change beliefs. We identify a key challenge to a constructive climate discourse by showing which political narratives are most likely to shape norms in real social media interactions: villain and human character messages are systematically more viral than those with instrument characters and hero roles. This dynamic may incentivize both proponents and opponents of climate change policies to use blame, especially of human characters, contributing to in- vs- out-group thinking and polarization instead of focusing on constructive, mutually beneficial reforms. We then show that assigning a drama triangle role to a character, while keeping objective embedded facts the same, can shape beliefs and memory and thus contribute to creating (or correcting) misperception. Overall our results help to better understand the reality of the climate change discourse, but also suggest that narrative-based interventions might be potentially promising compared to purely information-based approaches c.f. Haaland, Roth, and Wohlfart (2023).

Our work contributes to a large literature in media economics (see overview in Zhuravskaya, Petrova, and Enikolopov 2020) by providing a novel framework with application pipeline to analyze narratives in the media and by analyzing virality as an understudied outcome determining the effect of media on beliefs and preferences. Foundational work has examined newspapers (Gentzkow and Shapiro 2010; Djourelova, Durante, and Martin 2025), television (e.g., Ash et al. 2024a; Ash et al. 2024b; Ash and Galletta 2023; Ash and Poyker 2024; Enikolopov, Petrova, and Zhuravskaya 2011;

Durante, Pinotti, and Tesei 2019; Qian and Yanagizawa-Drott 2017) and radio (e.g., Yanagizawa-Drott 2014; Adena et al. 2021). An important focus has been biases in media reporting and their sources (Durante and Knight 2012; Cage et al. 2022; Caesmann et al. 2024), strategic decisions in the timing of publications (Durante and Zhuravskaya 2018; Djourelova and Durante 2022), and the impact of competition (Cagé, Hervé, and Viaud 2020; Cagé 2020). Related studies in political economy study folklore (Michalopoulos and Xue 2021) and movies (Michalopoulos and Rauh 2024), and narratives in a historical context Cagé et al. (2023). Closely related is (Bursztyn et al. 2023), who examine how exposure to different narrative on television can have downstream effects on beliefs and behavior

Finally, we contribute to the specific literature on the economics of social media (see overview in Aridor et al. 2024) by shifting the focus from platform exposure to the structure of the content that spreads. Prior work shows that social media can shape offline outcomes, including protest participation (Enikolopov, Makarin, and Petrova 2020) and hate crimes (Müller and Schwarz 2021; Müller and Schwarz 2023). Our approach complements this evidence in two ways. First, we provide a scalable, pre-defined content measure – character–role assignments – that predicts which political narratives become more viral on the platform, rather than proxying by topic keywords or sentiment alone (cf. Braghieri et al. 2024). Second, we connect those content primitives to persuasion outcomes in experiments, showing that single exposure shifts beliefs and revealed preferences about the characters embedded in the narrative, while leaving policy positions largely unchanged. Taken together, the findings help explain the persistence of certain political narratives in online dicourses: human characters (especially in villain roles) carry a systematic virality premium, which increases the likelihood of repeated exposure and thereby raises the potential for downstream behavioral impact.

2 Definition and Measurement of Political Narratives

2.1 Definition

We take a definition of political narratives that complements the "narratives-as-causal-sequences" approach in economics while returning to the broader spirit that treats narratives as communicative devices for focusing attention, encoding roles and identities, and shaping norms and behavior (Shiller 2017; Akerlof and Snower 2016). Formally, fix a topic T and a universe of characters $K = H \cup I$, partitioned into human characters H (individuals or collective actors such as corporations, parties, states, movements) and instrument characters I (policies, laws, technologies). For any text unit (tweet, paragraph, article), let $K' \subseteq K$ be the set of characters that appear. A role-assignment function

$$r : K' \to \{\text{hero}, \text{villain}, \text{victim}, \text{neutral}\}$$

maps each appearing character to either a drama-triangle role (Karpman 1968) or neutrality. We call the triplet (T, K', r) a political narrative if and only if there exists at least one $k \in K'$ such that

$r(k) \in \{\text{hero}, \text{villain}, \text{victim}\}$; if all characters are neutral, the text is about the topic but does not constitute a political narrative in our sense. This definition intentionally accommodates fragments and non-sequential formulations (e.g., "Corporations are villains") while remaining compatible with causal or temporal representations.

The Greta Thunberg passage introduced earlier provides a concrete illustration. The characters Energy Companies, Countless People, and Younger People/Activists are the nodes (K') in the graphs shown above. The DCG/DAG panels to the left depict possible causal linkages among these nodes. In real text, however, the direction of causal arrows is often ambiguous even when role assignment is clear. Our coding therefore starts from the nodes and their roles, which can be defined and measured directly; causal arrows, when identifiable, can then be added on top of this foundation.

There are many possible ways of condensing a large set of entities into a smaller set of roles, but the drama triangle roles are the smallest set that still preserves sufficient evaluative contrasts. It collapses many possible identities and roles into three broad archetypal roles, offering a parsimonious yet expressive partition of the identity space. Some alternative schemata like Aristotle's Poetics focus more on functional roles within a plot and with regard to story progression. Others feature larger sets of roles, ranging from 6-8 by Campbell, 7 in Propp's prominent morphology, to 12-16 archetypes according to Carl Jung. Many of these more distinct identities and roles can be meaningfully collapsed into the three drama triangle roles (see evidence in Bergstrand and Jasper 2018).[5] They provide a recurrent role schemata that is reflected in human story telling from biblical parables and classical epics to contemporary news and social media.

Table 1 clarifies how this role-based definition relates to causal-sequence approaches and illustrates various ways of assigning roles. Let $G = (K', E)$ denote a directed (a)cyclical graph over the appearing characters, where $E \subseteq K' \times K'$ collects causal or temporal arrows. A text qualifies as a causal narrative when $E \neq \varnothing$; it qualifies as a political narrative in our sense when $\exists\, k \in K'$ with $r(k) \in \{\text{hero}, \text{villain}, \text{victim}\}$. The rows of Table 1 then parse common cases. Statements like "A carbon tax is meant to raise the price of certain goods" satisfy $E \neq \varnothing$ but leave all $r(k) =$ neutral (causal only). "The carbon tax is stupid!" assigns an evaluative role to a policy instrument ($r(\text{Emission Pricing}) = \text{villain}$) without specifying arrows (political only). "Price increases due to the carbon tax raise the cost of living for Americans" both links nodes ($E \neq \varnothing$) and implicitly casts Emission Pricing as a villain to US People as a victim (both causal and political). Finally, "Tariffs are beautiful" is a fragment that assigns a (positive) stance but does not specify a sequence (political only).

Crucially, the transition from neutrality to a role can be signaled in multiple ways that our framework allows: via evaluative labels and attributions (e.g., "greedy," "reckless"), via agency and responsibility in the syntax (who helps whom, who harms whom), via contrasts in evaluation across

[5] Reducing roles to a simple positive–negative polarity erases a key evaluative contrast: the distinction between active agents and passive sufferers. In a sentence like "Corporations exploit Workers," both entities receive a negative valence if collapsed to "bad," yet the narrative logic depends on one being an active perpetrator and the other a passive target. Many alternative role systems explicitly include one or more victim roles (e.g. Propp and Jung).

Table 1: *Examples of Narratives*

Feature			Examples		Types of narrative	
Sequences (causal, temporal)	*Emotions*	*Character-Role(s)*	*General*	*Climate Policy*	*Causal Narratives*	*Political Narratives*
✓			Tariffs affect the terms of trade.	A carbon tax is meant to raise the price of certain goods	Yes	No
	✓		Recently, I became very curious regarding news about tariffs	Recently, I became very curious regarding news about the carbon tax	No	No
✓		✓	Tariffs on foreign competitors protect domestic producers	Price increases due to the carbon tax raise the costs of living for Americans	Yes	Yes
✓	✓	✓	Tariffs on greedy foreign competitors protect our struggling domestic producers	Price increases due to the stupid carbon tax raise the costs of living of vulnerable everyday Americans	Yes	Yes
	✓	✓	Tariffs are "beautiful"	The carbon tax is stupid!	No	Yes

characters within the same sentence, and – when present – via the position a character occupies in an explicit causal chain. Thus emotions and sentiment can *signal* roles, but they are neither necessary nor sufficient: a text can be emotionally flat and still cast a villain (e.g., "Firm X caused the spill"), or emotionally charged without assigning any character a role (e.g., "I'm anxious about climate change").

Narratives often appear as fragments – slogans or moral labels – rather than fully articulated stories. Formally, a fragment is a restriction of (T, K', r) to a smaller set of nodes $K'' \subseteq K'$ (and, where relevant, E to $K'' \times K''$). Such fragments typically sit inside grander, meta-narratives that readers implicitly know. In the Thunberg example, the grander narrative ties multiple characters across several sentences, whereas many real-world phrases will express only a piece of that structure (e.g., "Energy Companies profit while people suffer"). Our character–role coding captures both the frequent fragments we observe in corpora (right-hand panels in the DCG/DAG illustration) and the richer configurations when multiple roles co-occur (left-hand panels).

This representation also clarifies measurement strategy and the role of dimension reduction. In tools like topic models or RELATIO (Ash, Gauthier, and Widmer 2024), the pipeline begins by extracting many surface forms (entities, subject–verb–object triples) and then asks the researcher to aggregate ex post to a manageable set of entities. In our notation, this is a mapping from

a high-dimensional space into a lower-dimensional character set K (e.g., {*immigrants*, *migrants*, *refugees*, *asylum seekers*} $\mapsto$ IMMIGRANTS; {*carbon tax*, *emissions pricing*, *cap-and-trade*} $\mapsto$ EMISSION PRICING). We move this dimension reduction upfront: researchers define K ex ante, justify the choice, and then estimate $r(\cdot)$ for each $k \in K$ at the document level. This makes the object of interest – who is cast as hero, villain, victim, or neutral within topic T – transparent and portable across corpora. When desired, one can subsequently estimate E on the same K to recover causal structure; but identifying arrows always presupposes that the nodes have been defined.

Finally, the same logic is portable beyond climate-policy text. For monetary policy, central banks (FED, ECB) and their instruments (INTEREST RATES, QE) can be nodes in K and cast as heroes, villains, or victims depending on the narrative context. In immigration debates, IMMIGRANTS, NATIVE WORKERS, and POLICYMAKERS are natural human characters, while BORDER POLICY or SANCTUARY LAWS are instruments. In trade disputes, DOMESTIC PRODUCERS, FOREIGN COMPETITORS, and TARIFFS fill the same roles. Our framework keeps the unit of analysis – the character–role assignment – constant across settings, allowing researchers to compare narratives within and across topics and to layer causal arrows when the text, design, or auxiliary data support them.

2.2 Pipeline

Reproducibility and reuse. We release a Python package and example notebooks that (i) implement efficient batching logic and API calls, (ii) enforce simple consistency rules (e.g., one role per character), and (iii) produce a structured panel dataset with one row per document and columns for character presence and roles, suitable for statistical estimation. Our implementation is optimized for OpenAI (GPT-4o-mini) but is model-agnostic: any modern LLM that returns structured JSON can be used with minor modifications. The repository includes the final, pre-tested prompts and is available upon request; a fixed public link will be provided upon publication.

Objects and notation. Let $\mathcal{D}$ be the set of documents, $\mathcal{K}$ the set of user-defined characters for the topic, and $\mathcal{R} = \{\text{hero}, \text{villain}, \text{victim}\}$. The pipeline produces (i) a document-by-character presence matrix $\mathbf{M} = (m_{ik})_{i\in\mathcal{D},k\in\mathcal{K}}$ with $m_{ik} \in \{0,1\}$ and (ii) role indicators $\mathbf{R} = (r_{ikr})_{i\in\mathcal{D},k\in\mathcal{K},r\in\mathcal{R}}$ with $r_{ikr} \in \{0,1\}$ subject to a per-character exclusivity constraint $\sum_{r\in\mathcal{R}} r_{ikr} \leq 1$. If $m_{ik} = 1$ and $\sum_r r_{ikr} = 0$, the character is recorded as neutral. The stacked matrix $[\,\mathbf{M}\ \ \mathbf{R}\,]$ is the analysis-ready panel.

Five implementation steps. We summarize the workflow, visualized also in Figure 1; methodological details are deferred to the Appendix.

1. Select and define the topic. A precise topic definition anchors character selection and downstream analysis.

2. Identify the source and extract data. Common sources include digitized newspapers, social media, transcribed TV/radio/YouTube, and open-ended survey responses. Typical pre-processing includes language filtering, de-duplication, and optional geo-filtering. Full extraction details and data sources are reported in Subsection A.1 and Table A.1; geo-filtering is described in Subsection A.2.

3. Identify relevant characters. Character selection maps the topic into a manageable set of human and instrument nodes (cf. Section 2). Inputs can include literature review, exploratory tools (topic models, entity recognition, RELATIO), and domain reading. This step fixes the columns of $\mathbf{M}$ and the blocks of $\mathbf{R}$ that the model will fill.

4. Prepare the prompt(s). Prompt design specifies the mapping from raw text to $(\mathbf{M}, \mathbf{R})$. We use a two-stage structure: (i) a topic-specific relevance classifier (`irrelevant`/`assert`/`deny`/`policy`); (ii) conditional on `policy` (whether the tweets is about climate change policy or related), character presence and role assignment over $\mathcal{K} \times \mathcal{R}$. The package enforces basic consistency (e.g., one role per character). Final prompts are reproduced in Subsection A.3.

5. Obtain predictions and assemble outputs. We annotate documents via API (using batch submission for scale), parse JSON responses, and write a tidy panel with (i) Stage-1 flags, (ii) presence m_{ik}, and (iii) role dummies r_{ikr}. Figure 1 visualizes the classification flow. Operational annotation details are in Subsection A.3.

Quality control and validation. We recommend a light human audit and, where feasible, a small human-vs-LLM comparison. In our application, we conducted a 500-tweet MTurk exercise with two coders per tweet (28 workers in total). Our evaluation showed that agreement between GPT and humans on character presence and roles was comparable to, and in somec cases exceeding, inter-human agreement (see Appendix B).

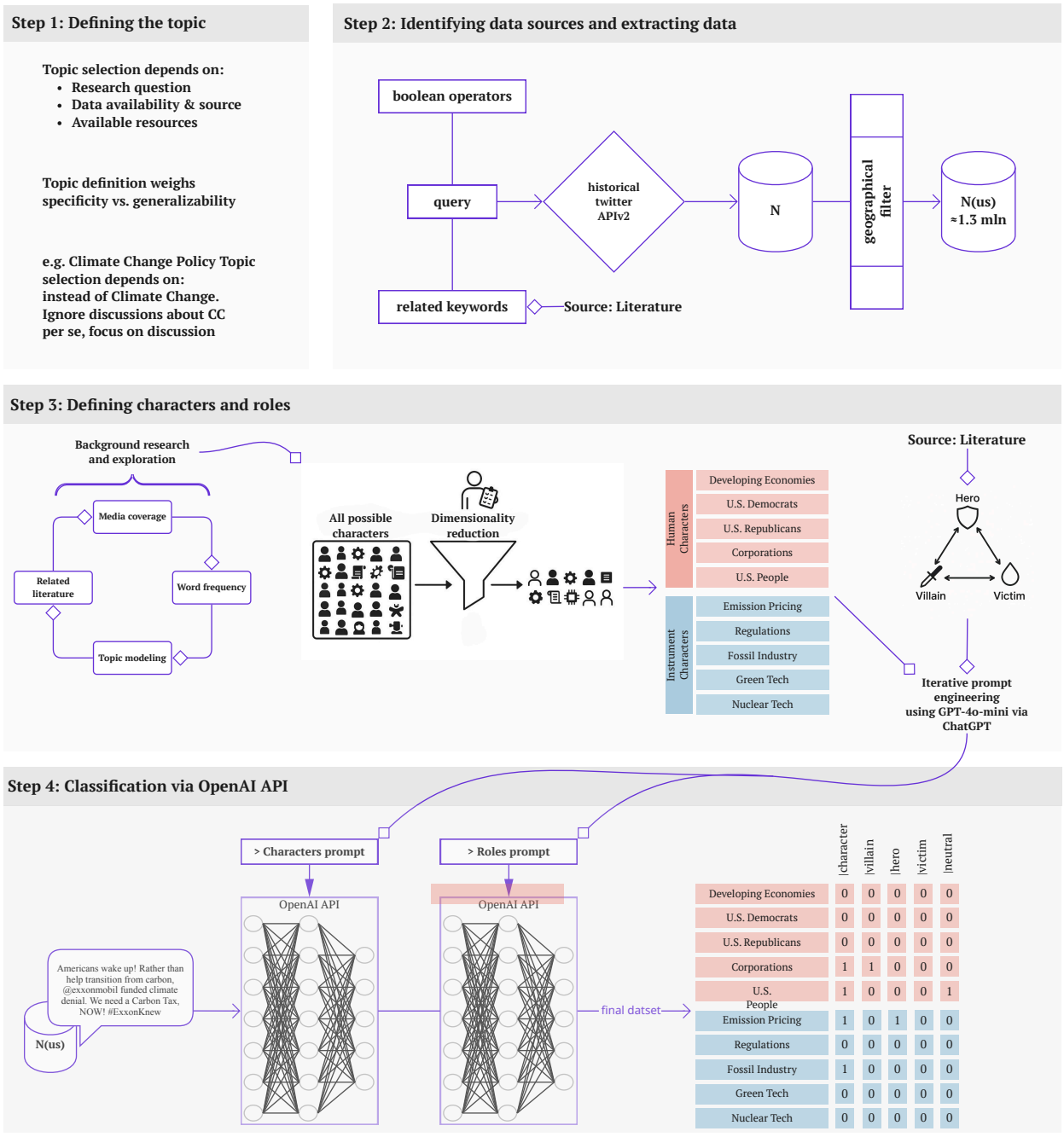
3 Data: Twitter Sample over 2010–2021

This section applies the five-step pipeline in §2.2 to our setting; methodological details are deferred to the Appendix.

1. Topic. We study political narratives about climate change policy (as distinct from climate change more broadly) in the United States during the 2010–2021 period.

2. Source and extraction. We query Twitter/X's historical APIv2 with a climate-policy keyword set adapted from Oehl, Schaffer, and Bernauer (2017), combining (i) one randomly selected day per month and (ii) every Saturday, to balance representativeness and tractability. We restrict to English, exclude retweets (retweet counts are used later as outcomes), and de-duplicate tweet IDs. We further

Figure 1: ***Visualization of the Classification Process***



Notes: The diagram illustrates Stage 1 (topic relevance) and Stage 2 (character presence and role assignment) for each tweet. The implementation parallelizes request preparation and uses API batch processing to scale to millions of tokens (see Subsection A.3).

validate language using `spaCy`'s language detection and normalize text (strip non-BMP unicode, standardize line breaks); details in Subsection A.1. We geo-filter tweets to the US using user profile locations and, when available, tweet geo-tags matched via `geopy`/Nominatim to OpenStreetMap polygons; see Subsection A.2. Data sources and availability are summarized in Table A.1. After

cleaning and geo-filtering, the analysis corpus contains 1,151,671 tweets.

3. Characters. Guided by the relevant literature, exploratory tools (word clouds, topic models, entity recognition, RELATIO), and in particular intensive domain reading (cf. Section 2), we pre-specify ten characters: five human – Developing Economies, US Democrats, US Republicans, Corporations, US People – and five instrument – Emission Pricing, Regulations, Fossil Industry, Green Tech, Nuclear Tech.

4. Prompts and annotation. We use a two-stage prompting scheme with GPT-4o-mini via the OpenAI API. Stage 1 flags topic relevance (`irrelevant`/`assert`/`deny`/`policy`). Conditional on `policy`, Stage 2 detects character presence and assigns at most one role (hero, villain, victim) per character, recording neutral otherwise. Prompt text and operational details are in Subsection A.3; the batch annotation flow is depicted in Figure 1. We use OpenAI's Batch modality to process the corpus efficiently.

5. Outputs and analysis sample. The annotation produces a tidy panel with (i) Stage-1 flags, (ii) character presence m_{ik}, and (iii) role indicators r_{ikr}. We define tweets as `relevant` if they concern the specified topic and comprise at least one of our pre-specified characters (neutral or in a role). This yields 309,744 `relevant` tweets in English language from the United States. Within `relevant` tweets, a tweet features a political narratives if $\exists\, k \in K'$ with $r(k) \in \{\text{hero}, \text{villain}, \text{victim}\}$; all other `relevant` tweets contain characters featured in a neutral way. Table C.3 reports detailed descriptive statistics, with mean words ≈ 29 excluding hashtags/mentions and mean retweets 3.7, with heavy-tailed dispersion.

Quality control and validation. We conducted a Mechanical Turk exercise on 500 randomly sampled relevant tweets, with two independent human coders per tweet using the same role taxonomy and instructions. Setup and agreement statistics are reported in Appendix B.

4 Descriptive Evidence

4.1 Frequency of character-roles

This section maps the landscape of political narratives in our corpus by describing the frequency of character–role assignments for all `relevant` tweets. Using the notation introduced in Subsection 2.2, let $m_{ik} \in \{0, 1\}$ indicate the presence of character $k \in \mathcal{K}$ in tweet $i \in \mathcal{D}$ and $r_{ikr} \in \{0, 1\}$ the assignment of role $r \in \mathcal{R} = \{\text{hero}, \text{villain}, \text{victim}\}$. Corpus composition is detailed in Figure C.2. Roughly 16% of climate-policy tweets mention no listed character, and we ignore those in our analysis. Our final sample of `relevant` tweets consist of those with characters, either depicted as neutral or cast in a drama triangle role. Table 2 summarizes character–role shares among `relevant` tweets. Panel (a) covers human characters and Panel (b) instrument characters.

Table 2: ***Share of Character–Roles in Relevant Tweets (United States, 2010–2021)***

Panel A: Human Characters

	Hero	Villain	Victim	Neutral	Total
Developing Economies	0.13	1.20	0.90	0.20	2.43
US Democrats	5.70	1.81	0.06	1.35	8.92
US Republicans	0.12	9.46	.	1.58	11.16
Corporations	0.98	8.14	0.06	7.87	17.05
US People	3.99	0.55	3.09	9.68	17.31

Panel B: Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emission Pricing	2.04	1.25	.	1.19	4.48
Regulations	3.48	1.36	.	3.18	8.01
Fossil Industry	0.06	11.75	0.04	1.60	13.46
Green Tech	11.52	0.84	.	3.19	15.55
Nuclear Tech	0.86	0.32	0.02	0.44	1.64

Notes: Shares are computed over the set of `relevant` tweets (those with $\sum_k m_{ik} \geq 1$). We report character–roles that appear at least 100 times over 2010–2021; excluded cells are shown as dots. "Neutral" indicates $m_{ik} = 1$ and $\sum_r r_{ikr} = 0$ for that character. Multiple character–roles may co-occur within a tweet. Appendix Table C.6 reports absolute counts (including categories excluded here).

Some patterns stand out. Among instrument characters, Green Tech as hero and the Fossil Industry as villain are the most frequent role assignments (both around 11–12% of all relevant tweets; Panel B). Second, among human characters, the US People is a prominent character appearing both as heroes (about 4%) and as victims (about 3%; Panel A). Corporations appear frequently as villains (about 8%). US Republicans as villains and US Democrats as heroes appear rather frequently, showing a clear role assignment within this policy space. We will not focus on parties as characters in this paper, but defer a more detailed analysis to future work. At the other end of the distribution, some role–character pairs are very rare or non-existent in this corpus (e.g., US Republicans–Victim, Emission Pricing–Victim, Regulations–Victim, Green Tech–Victim). Character-roles appearing fewer than 100 times over our sample period are excluded from further analysis. Table C.6 shows the absolute counts.

4.2 Features of Political Narratives

In this section, we explore the text features of political narratives. We create a variable *political narratives*, which takes on the value 1 if at least one of the characters are assigned to a drama triangle role, and the value 0 for the remaining tweets in `relevant` tweets with all neutral characters (for all $\sum_k m_{ik} \geq 1$). The aim is to descriptively explore whether narratives differ systematically in emotional content/valence and in language metrics, characteristics that might affect processing and sharing of the content. Emotional content/valence is not the only way of assigning roles, as discussed, but it is a plausible and common way of doing so.

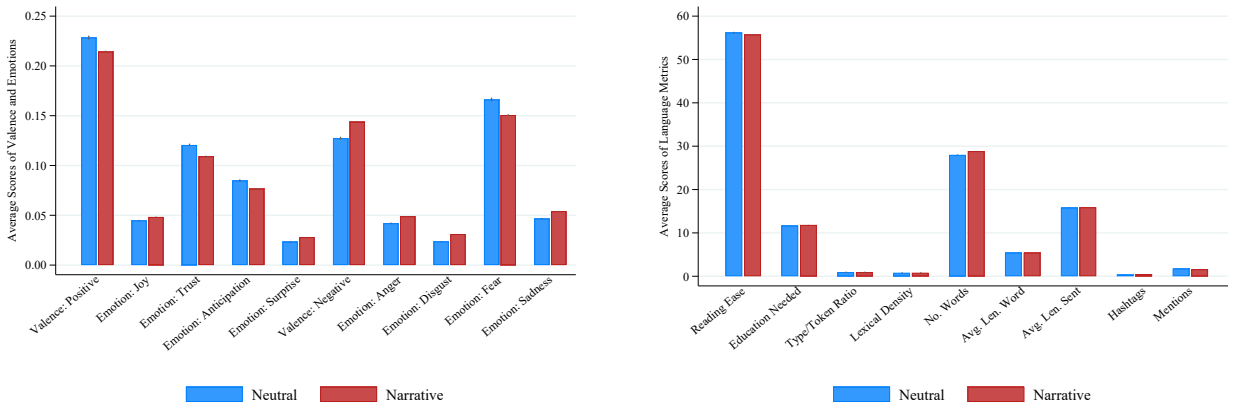
Figure 2 summarizes the comparison. The top row contrasts political narratives tweets, to

tweets where all characters are featured neutrally; the bottom row breaks out roles. Emotions and valence are computed using the NRC Emotion Lexicon (Mohammad and Turney 2013) (joy, trust, anticipation, surprise, anger, disgust, fear, sadness; valence is overall positivity/negativity). Language metrics include reading ease, years of education needed, type–token ratio, lexical density, number of words, average word length, and average sentence length (standardized for readability), as well as the number of hashtags ("#") and mentions ("@") as Twitter specific linguistic features.

Figure 2: ***Valence, Emotions, and Language Metrics of Relevant Tweets (United States, 2010–2021)***

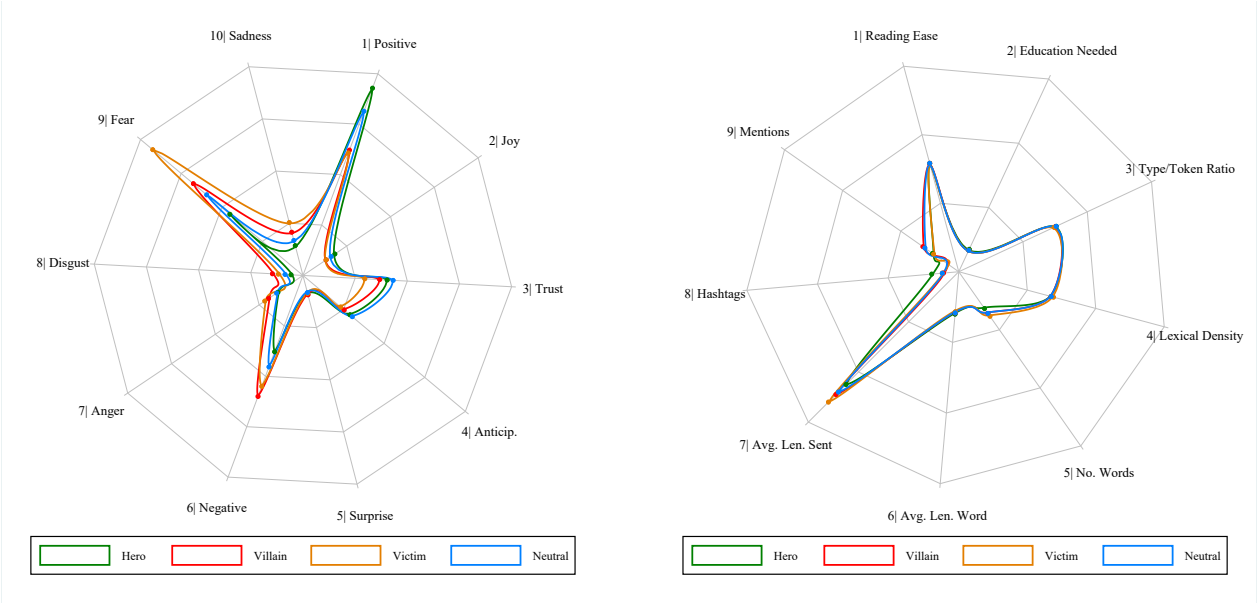
(a) Political Narrative vs. Neutral Characters: Emotions and Valence

(b) Political Narrative vs. Neutral Characters: Language Metrics



(c) Across Roles: Emotions and Valence

(d) Across Roles: Language Metrics



Notes: Panels 2a–2b compare tweets with characters all in a neutral role to political narrative tweets (at least one $r_{ikr} = 1$). Panels 2c–2d break out the latter by role. Scores are standardized for readability; higher values indicate greater presence of the corresponding emotion/metric. Emotion and valence measures use the NRC Emotion Lexicon (Mohammad and Turney 2013).

Four patterns emerge. First, political narratives exhibit more emotional content than their

neutral counterpart: when averaged across roles, they skew more negative in valence and load more on discrete emotions; this average masks clear heterogeneity by role. Second, the role breakdown is intuitive: hero narratives are the most positive; villain narratives are the most negative (with anger/disgust particularly salient); victim narratives load on fear. Third, language metrics are broadly similar across political narratives and neutral tweets; the main difference is length – political narratives are slightly longer on average, with victim narratives tending to be the longest. Fourth, the use of hashtags and mentions differs somehow between roles, but with no obvious patterns.

4.3 Political Narratives Over Time

Compared to the pure frequency table by roles before, Figure 3 presents the share and evolution of political narratives about climate policy over time. The 2010 to 2021 period features significant economic and political shifts, including general events like changes in presidencies, as well as topic-specific changes like the US leaving the Paris climate agreement. The figure focuses on the frequency of individual character-roles and does not yet account for their combinations within the same tweet. We organize the discussion using three ideas. A shift is any reallocation of attention across characters or across a character's roles; a reversal is when a character's or character types' dominant role flips; and we refer to a character as role-entrenched if its role hierarchy is stable over time.

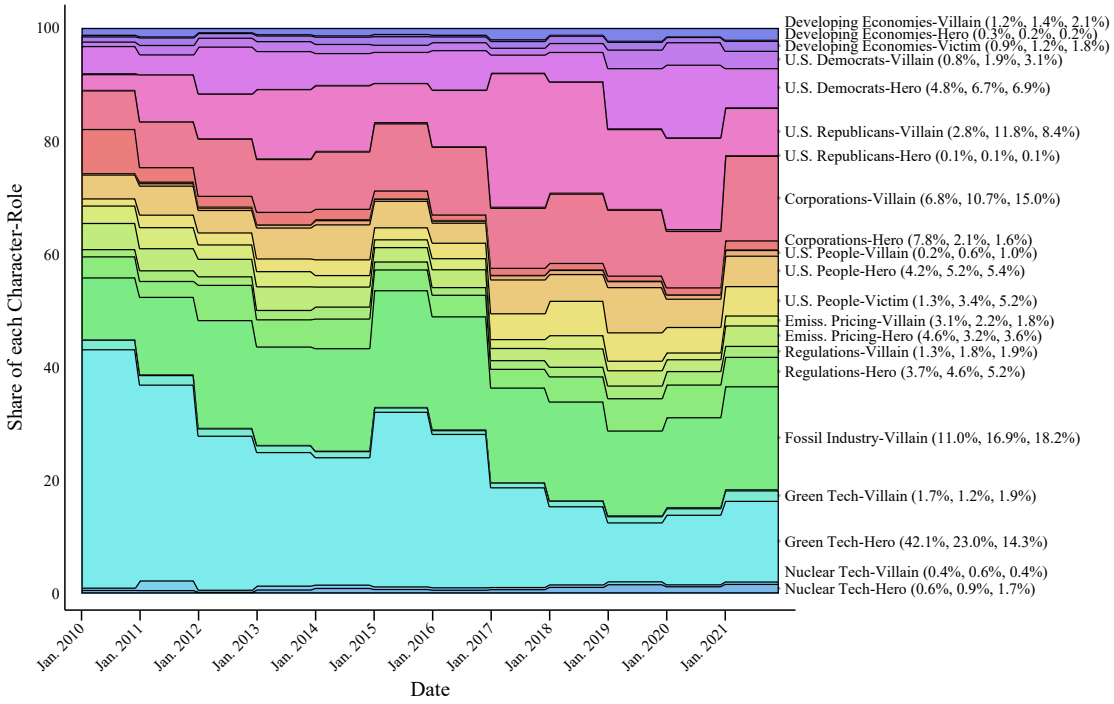
Let's start with entrenchment. Several characters exhibit stable distributions across roles. Green Tech exhibits the strongest domination by hero narratives; Regulation and Emission Pricing also tend to be dominated almost 3:1 (2:1) by hero roles. In contrast, Republicans are clearly dominated by villain, and Developing Economies slightly dominated by villain followed by victim.

We observe only few reversals alongside several notable shifts. Corporations are particularly striking. Corporations move from a slight hero edge (7.8% vs 6.8%) to a strong villain dominance (15.0% vs 1.6%). This dramatic change in the perceived role of private companies appears to be linked to two big shifts. Fossil Industry-Villain increases from 11 to 18.2%, while Green Tech-Hero declines from 42.1 to 14.3%. The only positive shift involving private companies is an increase in Nuclear Tech-Hero from 0.6 to 1.7%, which is large but not sufficient to overcome the overall negative shift in discourse about the role of private companies regarding climate change.

We can also zoom in on a specific instrument character, Emission Pricing, which comprises both the concepts of carbon taxes and emission trading. We observe a sharp decline in hero narratives by about a fifth, but villain narratives even drop by almost half. This seems to reflect that among those discussing emission pricing, including in the economics profession, opinions have clearly shifted in favor. However, despite the almost unanimous support of economists, the overall importance of emission pricing in the climate change policy discourse has clearly decreased.

Two descriptive insights stand out analytically. We observe (i.) a shift towards more human characters and (ii.) towards more villain roles. These shifts might be linked to the rise of populaism and Donald Trump, whose rhetoric heavily often dramatizes political discourse through the use of the drama triangle roles. The reallocation towards human characters at the expense of instrument

Figure 3: ***Frequency of Character-Roles over Time***



Notes: The figure plots annual shares of each character–role in the `relevant` tweets. For each year (2010–2021), the share of a character–role is its fraction among all identified character–role mentions that year. To maintain readability, we display labels for the 21 most frequent roles. Each label reports in order: start-year share, sample mean, end-year share. Unlabeled roles (top to bottom) are: US Democrats-Victim, Corporations-Victim, Fossil Industry-Hero, Fossil Industry-Victim, and Nuclear Tech-Victim.

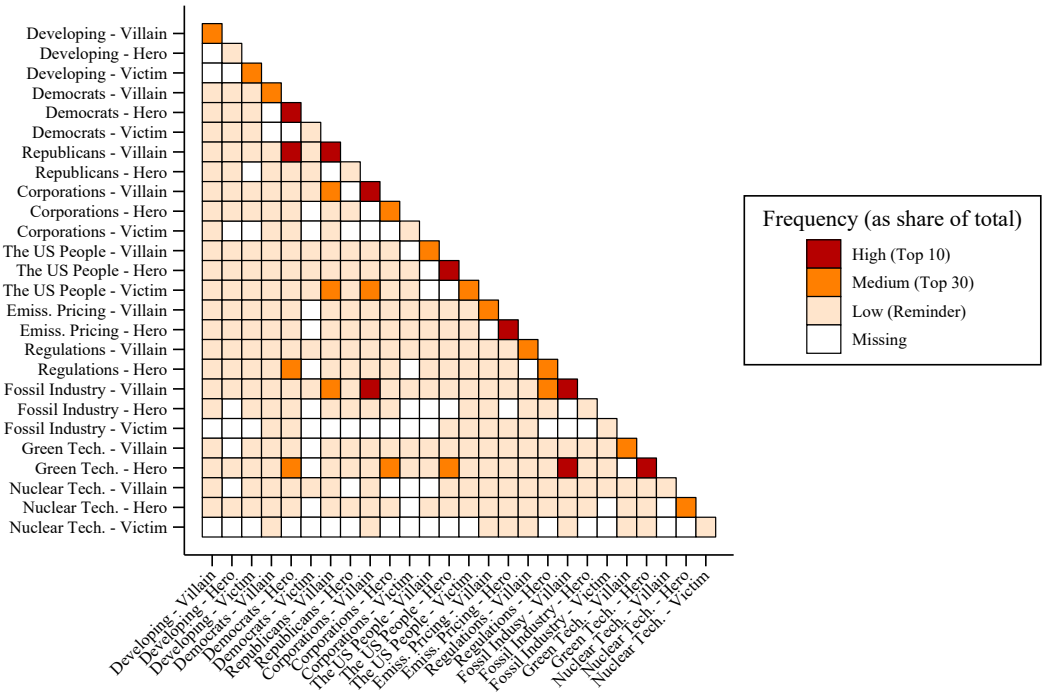
characters means there is less thinking and discussion about solutions, more about assigning merit and blame (institutions as hero characters drop by more than 25 percentage points). The shift towards villain roles signals both a growing frustration of all sides with climate change policies: for some it is way too little given the enormity of the challenge; for others even that little action is too much. This is visible in humans in villain-roles increasing by more than 15 percentage points and – while at a lower level – human victim roles also almost tripling. We refrain from strong causal claims with regard to the causes of these changes, which should be scrutinized in future research; the value here is to document where attention moved in the public climate change policy discourse on social media.

4.4 Character-role Combinations

The versatility of the political narrative framework allows us to move beyond marginal frequencies and investigate specific character–roles (CRs) combinations as a heatmap in Figure 4. This offers a more granular view of the narrative structures present in the corpus, and complements the time trends for individual CR shown in Figure 3.

To keep the heatmap interpretable, we restrict ourselves to the subset of `relevant` tweets with one or two character–roles, $\mathcal{D}^{\text{rel}}_{[1,2]} = \{\, i \in \mathcal{D}^{\text{rel}} : 1 \leq N_i \leq 2 \,\}$, with the number of (non-neutral) character–roles in tweet i as $N_i = \sum_{k \in \mathcal{K}} \sum_{r \in \mathcal{R}} r_{ikr}$. The diagonal elements depict single character-

Figure 4: ***Absolute Frequency of Character-Roles Combinations - Tweets with One or Two Character-Roles***



Notes: The figure shows the frequency of each character–role appearing alone or in combination with another character–role divided by all political narrative tweets with one or two character–roles. The diagonal of the matrix shows how often each character-role appears alone in a tweet. Tweets with three or more character-roles are excluded. We report character-roles that appear at least 100 times over 2010-2021; excluded characters are US Republicans-Victim, Emission Pricing-Victim, Regulations-Victim, and Green Tech-Victim. We avoid clutter, we use a color scheme to highlight the top 10 most frequent character-role combinations, the rest of the top 30, and the remaining pairs. White indicates a pair that never appears together. The top 10 are in order: Green Tech-Hero (10.27%), US Republicans-Villain (8.95%), Fossil Industry-Villain (5.56%), US Democrats-Hero (4.21%), Corporations-Villain (3.66%), US People-Hero (3.40%), Fossil Industry-Villain + Corporations-Villain (3.27%), Green Tech-Hero + Fossil Industry-Villain (3.02%), Emission Pricing-Hero (2.40%), US Republicans-Villain + US Democrats-Hero (1.92%). See Appendix Subsection C.3 for a more detailed explanation on the formulas used for computation, and Figure C.3 for a version of the heatmap including the numerical values of the top 20% most frequent shares.

role political narratives, and each off-diagonal element a combination of the CR with another CR. For simplicity, we focus on highlighting the top 10 and top 30 most frequent combinations in different colors. To compute the shares for each heatmap element, we divide the absolute number of times the single CR or CR-combination appears by the total number of tweets included in $\mathcal{D}^{\text{rel}}_{[1,2]}$.

This approach allows us to both identify the most frequent combinations and to contrast single appearance with combinations for each CR. If the diagonal element is the most frequent, this CR most often appears as a stand-alone narrative fragment. If a single or the sum of off-diagonal elements are more frequent, this indicates that complex narratives are more important in shaping the perception of this CR on social media.

For all character-roles, the simple narrative form is the most common way in which they

appear. In line with the prior graph on narratives over time, we find that besides political parties Corporations-Villain, Fossil Industry-Villain and Green Tech-Hero are the most frequent political narratives. This exemplifies how important it is in a real text corpus to also capture these narrative fragments, even though they might not contain a causal relationship between characters.

Combinations of CRs are also important with regard to frequency, with a few specific CRs and combination standing out. For instance, the combination Corporations-Villain with Fossil Industry-Villain is in the top 10 most frequent, as is the antagonistic combination Green Tech-Hero and Fossil Industry-Villain. The hero-hero combination Green Tech-Hero with US People-Hero is also among the top 30. We can also observe that some CRs never occur together as a combination, for instance the combination Fossil Industry-Hero and Emission Pricing-Hero. To put these numbers in perspective, out of the total number of `relevant` tweets $\mathcal{D}^{\text{rel}}$, the share of single CR political narratives is approximately 46%, 24% are two CR-combinations, and 12%contain three or more CRs. The remaining 18% of tweets feature characters only in a neutral way (see Subsection C.3).

5 Main Results: Virality of Political Narratives

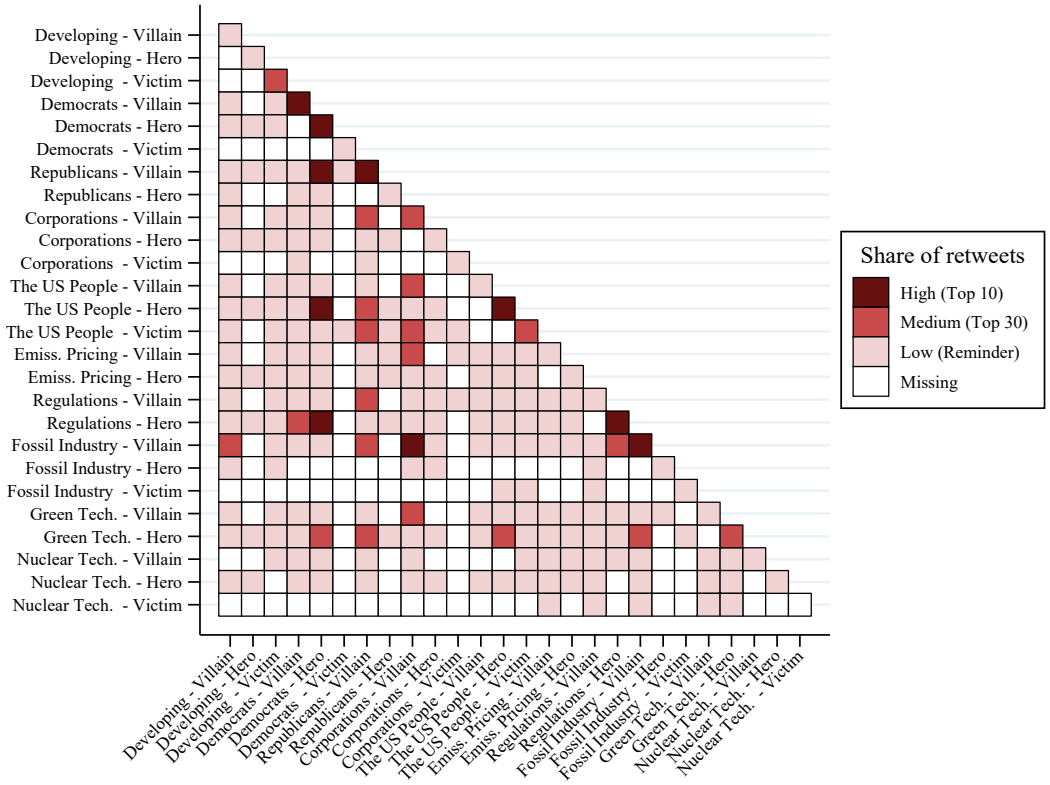
In this section, we begin by examining our main outcome variable, the number of retweets. We briefly discuss a heatmap of the most viral character-role combinations and analyze the shape of the retweet distribution for political narratives versus those tweets with all neutral characters. Based on this examination, we then turn to our systematic regression analysis of the determinants of virality. We examine whether political narratives are on average more viral than neutral tweets with the same characters, and then turn to a more detailed analysis of roles and role combinations. In the final section, we examine possible heterogeneity between human and instrument character types, as well as possible benefits of combining several character-roles in one tweet.

5.1 Virality of Character-Roles

Before turning to the main empirical results, we begin by examining variation in narrative virality across all one- and two-role combinations. Figure 5 presents a heatmap based on the retweet shares of each combination, again focusing on the top 10 and top 30 combinations. We compute retweet shares as the sum of all retweets with a particular CR-combination over the aggregate number of retweets of all combinations. Computing the average number of retweets per cell risks being misleading, as a single tweet with an unusual retweet number biases the computation of the average. As before, we can both learn about the absolutely highest shares, but can again also compare CR-combinations with single-CR narratives. While some patterns mirror those in the frequency matrix, some notable divergences emerge. This highlights that the character–role combinations most frequently composed by users are not necessarily the ones that receive the most retweets.

Certain character-roles and CR-combinations turn out to be retweeted particularly often. Fossil Industry-Villain appears in four different combinations and as a single fragment in the Top 30,

Figure 5: ***Virality of Character-Roles - Tweets with One or Two Character-Roles***



Notes: The figure shows the retweet share of each character–role appearing alone or in combination with another role, among relevant tweets containing one or two roles. Retweet rates are computed as the share of total retweets received by a given role (or pair) relative to all retweets of tweets with one or two roles. The diagonal of the matrix shows the retweet rate when each character-role appears alone. Tweets with three or more character-roles are excluded. We report character-roles that appear at least 100 times over 2010-2021; excluded characters are US Republicans-Victim, Emission Pricing-Victim, Regulations-Victim, and Green Tech-Victim. To avoid visual overload, we do not display exact rates. Instead, we use a color scheme to highlight the top 10 most frequently retweeted character-role combinations, the top 30 (which includes the top 10), and the remaining pairs. White indicates a pair that never appears together. The top 10 in order is: US Republicans-Villain (16.72%), US Democrats-Hero (12.98%), US People-Hero (7.93%), Fossil Industry-Villain + Corporations-Villain (7.89%), US Republicans-Villain + US Democrats-Hero (6.51%), US Democrats-Villain (4.80%), Fossil Industry-Villain (3.08%), US People-Hero + US Democrats-Hero (2.71%), Regulations-Hero (2.58%), Regulations-Hero + US Democrats-Hero (2.52%). See Appendix Subsection C.3 for a more detailed explanation on the formulas used for computation, and Figure C.4 for a version of the heatmap including the numerical values of the top 20% most frequent shares.

only beaten by eight top 30 elements that linked to Republican-Villain. Regarding specific combinations, Fossil Industry-Villain with Corporations-Villain is a villain-villain combination featured in the top 5 highest retweet shares. Beyond that, the frequent combinations US People-Hero & Green Tech-Hero as well as the antagonistic combination Green Tech-Hero vs. Fossil Industry-Villain are also among the top 30 most retweeted. The US People character appears overall seven times in the top 30, three times as hero, three times as victim, but also once as villain in combination with Corporations-Villain.

In many cases, these combinations allow inference about the structure and possible causal logic of the underlying narratives. Regulation-Hero is an interesting example. It appears both

in the top 10 in combination with DEMOCRATS-Hero as well as in the top 30 with DEMOCRATS-Villain. This may be seem surprising at first glance, but could reflect different expectations among social media user groups. The narratives of one group highlight Democrats' active role in the fight against climate change by drafting and enacting important regulation. The other group shares the enthusiasm for regulation as a tool, however, blames them for a lack of engagement regarding more and better regulation. Manual inspection of tweet examples confirms that many tweets containing these character-role combinations follow such a pattern.

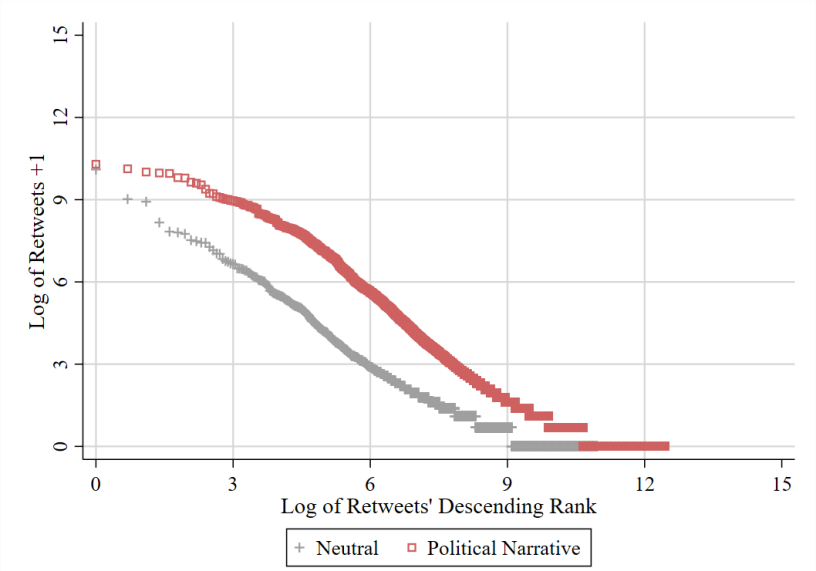
To move beyond individual character–roles, we then begin examining broader narrative patterns by shifting the focus to the virality of political narratives. This involves aggregating across roles, combinations of roles, and comparing narrative types based on whether they involve human or instrument characters. We begin with a fundamental question: Are political narratives more viral than comparable tweets that discuss the same topic and characters but contain no drama triangle roles? While it is theoretically possible that some roles are less viral than neutral framings, and others more so, prior qualitative work and research from related fields lead us to expect that political narratives tend to generate greater virality on average. We therefore start by comparing the virality of tweets that contain political narratives to those that include characters but assign no role – that is, neutral tweets.

Comparing the unconditional distributions reveals that political narratives are generally more viral than neutral tweets, with the difference being larger for the upper end of the retweets counts. Figure 6 displays the Log-Log Rank Distribution of retweets for both categories as the most suitable way for a comparison given the underlying very-skewed distributions. The x-axis displays the logarithm of the tweet rank based on their retweets, with rank 1 corresponding to the most retweeted tweet in the dataset, and the y-axis the logarithm of the number of retweets (plus 1). The vertical position at a given rank reflects relative virality: a higher curve means tweets in that category at that rank receive more engagement than tweets at the same rank in a category with a lower curve. We find that the political narratives curve is always higher than the neutral tweets curve, and that the vertical distance between the curves tend to be larger the higher the rank of a tweet within the respective distribution.

While this comparison remains descriptive, we can derive some interesting metrics from the comparison[6]. At the lower end of the rank distribution, political narrative tweets are 1.6 times more likely than neutral ones to exceed 100 retweets. This distance increases to 1.9 (2.6) times as likely for exceeding 500 (1000) retweets. Of course, these differences might still reflect many unobservable characteristics of tweets and other unobserved factors. For that reason, we now shift to a more systematic regression analysis, which tries to carefully dissect potential threats to identification and move closer to a potentially causal interpretation.

[6]We compute nonparametric survivor (complementary CDF) probabilities for retweets, $S(x) = \Pr(R \geq x)$, separately for tweets with any character-role narrative vs. neutral role. We then report tail risk ratios $S_{\text{Narrative}}(t)/S_{\text{Neutral}}(t)$ at thresholds $t \in \{50, 100, 500, 1000\}$.

Figure 6: ***Log-Log Rank Distribution of Retweets in Relevant Tweets (United States, 2010-2021)***



Notes: The figure shows the log–log rank distribution of retweets for `relevant` tweets, distinguishing between those with at least one character–role (red) and those with characters only in a neutral form (gray). The x-axis plots the logarithm of the rank, with rank 1 as the most retweeted tweet in the sample. The y-axis plots the logarithm of retweet counts (plus one). The slope of the curve indicates how quickly the distribution falls from the most viral to the least viral tweets: a steeper slope means engagement is clustered in a handful of viral tweets, whereas a flatter slope indicates that engagement is more evenly distributed. The vertical position at a given rank reflects relative virality: a higher curve means tweets at that rank receive more engagement than tweets at the same rank in a dataset with a lower curve.

5.2 The Determinants of Virality

We turn to a regression framework to analyze the determinants of virality more systematically. Given the heavy-tailed distribution of retweets documented above, we estimate PPML models, which naturally accounts for zeros in count data and handle heteroskedasticity without log-transforming the dependent variable. We estimate the following regression equation:

$$E[Y_{i,s,t}|P_{i,s,t}] = exp[\alpha + \beta P_{i,s,t} + \theta_1 T_{i,t} + \delta_{s\times t} + \gamma_{s\times w(i)} + \eta_{h(i)} + \sum_{k\in\mathcal{K}} \eta_k \, c_{ik}] \tag{1}$$

where $Y_{i,s,t}$ refers to the count of retweets for tweet i originating from state s in week t. α is the constant term and $\delta_{s\times t}$ refers to a year $\times$ state fixed effect.[7] We include calendar-week times year FE $\gamma_{s\times w(i)}$ to absorb news and platform shocks specific to the time within a year, as well as hour-of-the-day FE $\eta_{h(i)}$ to capture diurnal cycles (systematic within-day patterns in Twitter engagement). $\sum_{k\in\mathcal{K}} \eta_k \, c_{ik}$ are character dummies where $c_{i,k} = 1$ indicates if character k appears in tweet i, so that identification comes from within-character variation in role assignment. $T_{i,t}$ is a vector of control variables at the tweet level, including the language metrics and measures of valence and emotions

[7] We include as an additional state the 'USA', for those tweets that we could only locate at the national level and not at the state level.

introduced in Subsection 4.2.[8]

$P_{i,s,t}$ refers to our variable of interest. In our first main test, P takes on the value 1 if the tweet contains a political narrative, and 0 otherwise. Our sample is $\mathcal{D}^{\text{rel}}$, i.e. the tweets with $P_{i,s,t} = 0$ are also on the same topic and contain at least one character. β is our coefficient of interest capturing the relationship between political narratives and virality. In our main specification, we cluster standard errors at the week level, and show later that our results are robust to other choices.

Figure 7 displays a comprehensive coefficient plot with six estimates ranging from the most simple to the most restrictive specification. The first row shows the simple correlation without any controls. Political narratives seem to be clearly more viral. The unconditional coefficient is large, positive, and statistically significant at the 1% level. The magnitude of the coefficient is also large. Coefficients are interpreted as percentage changes using the transformation $e^{\beta} - 1$. A political narrative does on average generate 80% more retweets than a comparable tweet with only neutral characters.

Our first concerns are temporal or spatial shocks that overlap with the treatment and outcome. National and local news, disasters, policy announcements and seasons could shift both role use and engagement. In our specification, the calendar–week times year fixed effects absorb national as well as platform shocks specific to the time within a year, and also control for time trends and seasonality in the most flexible way possible in our setting. Hour-of-the-day FE capture diurnal cycles, referring to systematic within-day patterns engagement in role use and virality on Twitter. Finally, state times year fixed effects allows both for heterogeneous effects of temporal shocks across states, and absorbs any state-year specific natural or policy-related events. Adding this very comprehensive set of fixed effects actually slightly increases the point estimate to 0.707, corresponding to 102.8% percent.

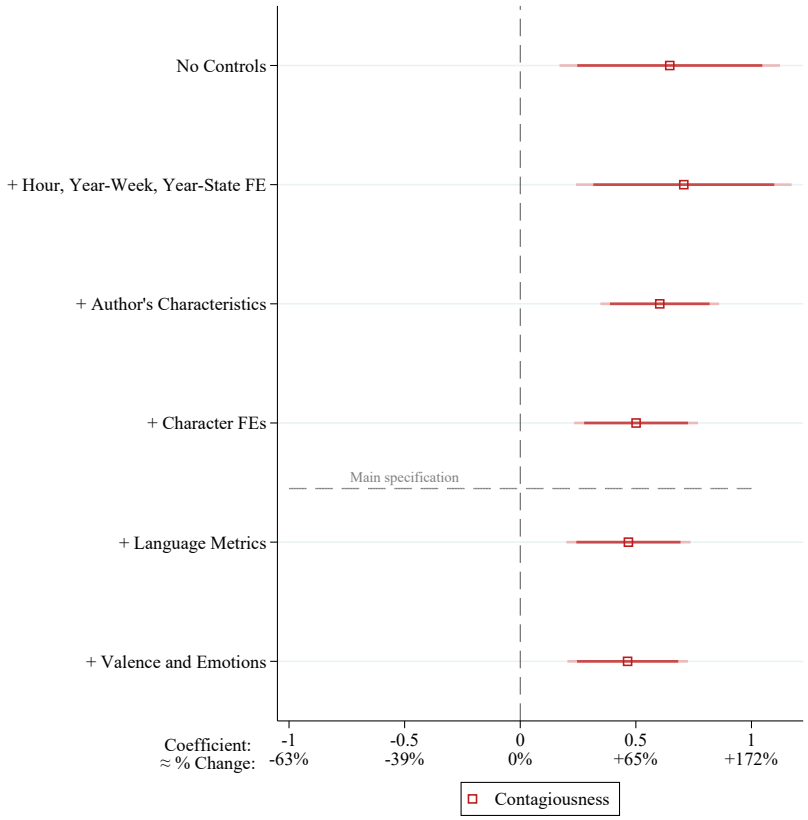
A second concern relates to author characteristics, which might correlate with both the usage of political narratives and virality. The most obvious difference is the number of followers: an account with many followers can go viral much more easily than a small account. If large accounts also tend to use more political narrative elements, this could lead to a spurious positive correlation between *political narrative* and retweets. By controlling for all available authors characteristics, we only evaluate political narratives against neutral tweets from comparable authors. Controlling for author characteristics attenuates the point estimate from 0.707 to 0.603 (or by 20.1%), while remaining clearly statistically significant. [9]

A third concern is that certain characters could be both more likely to be cast in drama triangle roles and generally tend go more viral, for reasons unrelated to the use of political narratives. This would also lead to a spurious positive correlation and an upward bias in our estimate. Controlling for character FE eliminates this concern, by only using variation within characters, e.g. comparing tweets about green technology in a drama triangle role with those where it is portrayed neutrally.

[8] There might also be measurement error due to misclassifications of the role labels $D_{i,s,t}$. If we assume that this is approximately mean-zero and independent of $Y_{i,s,t}$ conditional on covariates it should act like classical measurement error, attenuating estimates toward zero under a linear approximation

[9] Figure D.2 show no general trends visible in terms of correlations between categories of users by number of followers and frequency of character-role use.

Figure 7: ***Regression Results - Impact of Political Narratives on Virality***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regression models testing the effect of featuring at least one character-role vs. featuring characters only in a neutral role on virality, measured as the count of retweets. The x-axis reports coefficient estimates along with the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. Panels display results from increasingly restrictive models. Author characteristics include: verified status, number of followers/followings, total tweets created, party affiliation, religiosity, higher education, and parenthood. Language metrics and valence/emotions are defined as described in Subsection 4.2. Table D.1 shows a table with the full regression results.

Implementing this does shift the coefficient considerably from 0.603 to 0.501, however it remains large and clearly statistically significant with a p-value ≤ 0.001. We refer to the comprehensive set of controls in this specification from now on as our "main specification."

In our main specification, we deliberately do not control for language metrics and emotions/valence. Language metrics (readability/complexity measures) and emotions/valence are natural candidates to explain engagement, but may lie on the causal path from assigning drama–triangle roles to retweet engagement. For example, if authors frame CORPORATIONS as villain, they typically choose stronger verbs, moral language, and emotional words, which in turn affect virality. Similarly, we found that hero-narratives use more hashtags, shorter sentences and fewer words, also possibly related to virality. Because these features may also proxy inputs to platform ranking systems, including them can reduce confounding from algorithmic surfacing, but comes at the cost of conditioning on mediators (bad controls). Hence, including them could move us closer to the causal effect or to better understand the roles of these features as mechanisms. However, conditioning on bad controls also

introduces a potential bias in either direction. Hence, the coefficient β from the main specification remains our preferred estimate of the size of the total effect of political narratives on virality.[10]

Adding language metrics and valence & emotion proxies remains an imperfect but informative test; however, it turns out to have surprisingly limited influence on the point estimate. The last two estimates in Figure 7 re–estimate the main specification by adding the controls for language features, and then the ones for emotion and valence. The point estimate becomes slightly smaller (the percentage changes moving from 65% to 59.7% and 59.0%), and remains positive and statistically significant with a p-value ≤ 0.001. We interpret this as a strong indication that the predictive content of the political narrative framework is far from fully captured by simple langue metrics, emotions or valence.

We begin with a brief evaluation of robustness, since all subsequent analyses build on this core result. Readers who want to skip the next short robustness section can also directly proceed to our analysis of individual roles, role combinations and heterogeneous effects by character type and number of character roles.

5.2.1 Robustness Checks and Additional Results

Before moving to roles and role combinations, we conduct a series of robustness checks and a short evaluation of treatment heterogeneity to assess the stability and validity of this main result. In this section, we describe the main tests and their motivation, linking to the respective appendix section containing more detailed tables or figures.

The first and most important consideration is the sensitivity of the result to the algorithm, which could bias the coefficient downwards or, more problematically in our case, upwards. Without knowing details of the algorithm during our sample period, the best we are able to do is control for text features that might positively correlate with political narratives and drive virality. Another way to test for the potential sensitivity of our results towards the algorithm is to examine the effect of important changes in the algorithm on that result. If the result would be driven to an important part by the algorithm, we would expect that critical changes in it also affect the result. In contrast, if the relationship between political narratives and virality is largely independent of the algorithm, we would expect also little impact of algorithm changes.

We examine the introduction of an algorithmically ranked timeline in February 2016 as the arguably most important change in the Twitter algorithm during our sample period. In Subsection D.3, we use both a split-sample approach as well as interaction models to test whether the association between narratives and virality differs before and after this change. We find that the coefficient is positive and statistically significant before and after the change in the split-sample regressions, and basically indistinguishable in size. The interaction models yields an interaction coefficient that is close to zero and far from conventional levels of statistical significance. Together with the earlier results concerning the robustness to adding sets of extensive language metrics and

[10] Subsection D.1 provides a simple causal graph clarifying when language and emotions (or language metrics) act as mediators versus confounders, and why controlling for them as bad controls can create a bias.

emotion measures, this provides further assurance that the positive effect of using political narratives on virality is not a mechanical result of algorithmic timeline curation, but reflects a deeper and more fundamental relationship.

A second concern on social media are bots. It is possible that bots retweet political narratives more often. While this would still mean political narratives were factually more viral on social media during our sample, it would also indicate the result might be driven by non-human behavior. There is no perfectly reliable way to identify bots – or they would not be such a challenge – so we follow the computer science literature and proxy for bot origin based on high posting volume, low engagement, and frequently duplicate text-posting (or alternatively based on text repetition). Subsection D.4 shows that a conservative approach dropping all suspicious tweets changes neither size nor significance of our main results. Similar to the algorithm those tests are not definitive proof, but the lack of sensitivity to these changes reassures us that such a mechanism is not driving the results.

Third, Subsection D.7 and Subsection D.8 examine alternative, although less direct, proxies for virality. We show that using likes, the obvious other main "currency" on social media, as an outcome produces very similar results. Once author characteristics are controlled for, the generally positive association with the number of replies also turns clearly statistically significant. These alternative results can either be seen as additional evidence for the virality of political narratives, or alternatively as additional evidence of the broader impact of using such narratives beyond a narrow definition of virality.

Fourth, we also show that are main result is robust with regard to sign and significance when using alternative transformations of retweets as the outcome. Subsection D.6 shows that our results are robust to using OLS with a log or asinh transformation, as well as with a piecewise log-link specification to separately examine extensive and intensive margins (see Chen and Roth (2024). All alternatives yield results consistent with the main Poisson model, reassuring us that our results are not specific to a particular functional form. Subsection D.5 also validates that statistical inference is robust to alternative clustering levels, including state, month, and state–week.

Fifth, we address a potential concern that political narrative use or its effect on virality might differ across user types. This would not invalidate the main result, but if political narratives would only explain a higher virality of very small or very large account, this would reduce the external validity of our result. In Subsection D.2, we divide users into three groups by follower count to proxy for their typical reach: low (0–1,000), medium (1,001–10,000), and high (10,000+). If high-reach users systematically employ different types of narratives—or if narrative effects only materialize for already-prominent users—our results would be limited in scope. However, we find that users across all reach levels use remarkably similar narrative structures. Furthermore, the virality advantage of narrative tweets persists within each user tier, with slightly larger effect sizes among medium and high users.

Table 3: ***Regression Results - Impact of Individual Roles on Virality***

Dependent Variable	**Retweets' Count**						
	Coeff./SE/p-value						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hero	0.445			0.412		-0.552	0.239
	(0.193)			(0.211)		(0.210)	(0.279)
	[0.021]			[0.051]		[0.009]	[0.393]
Villain		0.923		0.964	0.552		0.791
		(0.135)		(0.198)	(0.210)		(0.291)
		[0.000]		[0.000]	[0.009]		[0.007]
Victim			0.090	0.173	-0.239	-0.791	
			(0.326)	(0.287)	(0.279)	(0.291)	
			[0.783]	[0.546]	[0.393]	[0.007]	
Neutral					-0.412	-0.964	-0.173
					(0.211)	(0.198)	(0.287)
					[0.051]	[0.000]	[0.546]
Sample: Neutral	✓	✓	✓	✓	✓	✓	✓
Sample: Hero	✓			✓	✓	✓	✓
Sample: Villain		✓		✓	✓	✓	✓
Sample: Victim			✓	✓	✓	✓	✓
Percentage Change	56%	151.7%	9.4%				
Mean Outcome (Control Group)	1.97	1.97	1.97	1.97	4.38	3.70	2.40
Pseudo R2	0.79	0.70	0.68	0.71	0.71	0.71	0.71
Observations	77047	90057	43952	137664	137664	137664	137664

Notes: The table reports coefficients from Poisson Pseudo-Maximum Likelihood regression models with retweet counts, our virality proxy, as the dependent variable. The sample includes `relevant` tweets with up to one CR, categorized into mutually exclusive categories (e.g., only hero or only neutral). Column 1-3 always compare on drama triangle role to neutral tweets, columns 4–7 each jointly includes all tweets and categories from column 1 to 3 and vary the reference category. Coefficients can be transformed into percentage changes using the transformation $e^{\beta} - 1$. All regressions include hour-of-the-day, calendar-week-times-year, and year-times-state fixed effects, author characteristics, as well as character fixed effects. Standard errors are clustered at the week level.

5.2.2 Virality of drama triangle roles

We now move on to test the individual effectiveness of the drama triangle roles, and show that some roles are more effective than others. The first three columns of Table 3 present the results of pair-wise regressions testing the effect of containing a specific role on a tweet's virality, always compared to tweets with only neutral characters. To begin with the cleanest and most transparent comparison, we only include tweets with exactly one character in these analyses (corresponding to the diagonal elements of the matrices we discussed before). Column 1 shows that hero narratives are statistically significantly more likely to go viral, with on average 55% percent more retweets. Villain narratives, in column 2, are also significantly more viral, and even associated with on average 150% more retweets. Tweets containing only victim narratives, however, are almost indistinguishable from neutral tweets (column 3).

Both hero and villain narratives turn out to be more viral than neutral tweets, and villain narratives turn out to be significantly more viral than any other role. Columns 4-7 of Table 3

test for significant differences across roles when putting all tweets types in a joint regression, each leaving out a different role as a reference category to transparently display all individual differences and their statistical significance. The key insight in column 4 is the confirmation that both hero and villain narratives are associated with significantly more retweets than neutral also in this more comprehensive comparison. Columns 5 to 7 confirm in particular that villain narratives are significantly more viral than hero and victim narratives, and that pure victim narratives are not associated with a statistically significant virality advantage in any specification. While these are important insights for tweets with one role, we continue in the next section by shifting attention to tweets with multiple character-role combinations and differentiate between human and instrument character types.

5.2.3 Virality of Role Combinations

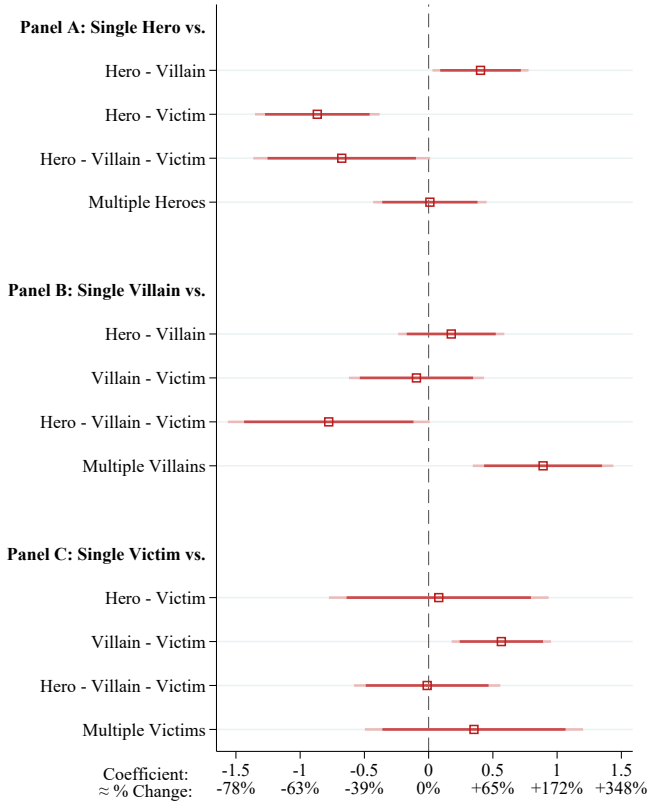
So far, our analysis has focused on the determinants of virality, examining both the overall impact of narratives and the effects of political narratives with different roles and combinations of character-roles. In this section, we present the final results of our empirical analysis, shifting the attention to the role of multiple character-roles and different types of characters.

We now move from the analysis of narrative fragments involving only a single character-role to a systematic analysis of role combinations, by extending the sample to all tweets containing political narratives. While narrative fragments are overall more frequent on social media, Figure 5 showed that they are not necessarily more viral than specific role combinations. Our systematic analysis of role combinations here builds on the prior analysis of individual roles and examines whether adding different roles or more characters cast in the same role further amplifies or weakens the individual role's effect on virality. For that purpose we run additional regressions where the role combinations are always compared to the respective pure role tweets as a reference category.

The results of adding additional roles to single hero, villain or victim narratives in Figure 8 reveal three main patterns. First, adding an additional villain character always further boosts virality, even if it means adding a second villain to an already existing villain character (multiple villains). Second, adding an additional hero character never has any significant effect on virality, and adding a victim even lowers the virality in one case. Thirdly, complex narratives with characters in each drama triangle role are significantly less viral than single hero or single villain narratives. This reinforces the apparent power of using villain roles to boost virality.

Several caveats qualify the interpretation of these results. First, we are not actually measuring the effect of adding those roles, but rather comparing tweets with single to those with multiple roles. Second, we have relatively few victim characters in our set of characters, which is reflected in the larger uncertainty surrounding the victim role results. Thirdly, more complex narratives with combinations of all drama triangle roles might struggle to go viral in the limited and contested attention space on social media. Other media with less space restrictions might naturally be more suitable to feature and share those grander narrative. Grander and more complex narratives might not be the most viral directly on social media, but they may serve as the basis and underlying

Figure 8: ***Regression Results - Heterogeneity in the Impact of Character-Role Combinations on Virality***



Notes: The figure shows coefficients from Poisson Pseudo-Maximum Likelihood regression models testing the effect of character-role combinations on virality, measured as the tweet-level count of retweets. The x-axis reports coefficient estimates together with the corresponding approximate percentage change, computed as $e^{\beta}-1$. Each panel corresponds to a separate regression model. Panel A examines the effects of hero combinations, with the comparison group being tweets featuring a single hero character only. The regressors capture cases where a hero is paired with one or more villains, with one or more victims, with both villains and victims, or with additional heroes. Panel B follows the same structure for villains, with the reference group being tweets that feature a single villain character, while Panel C does so for victims, using tweets with a single victim character as the baseline. All regressions include hour, week of the year, and year–state fixed effects. Standard errors are clustered at the week level, covering all weeks in the study period.

reference for many simpler narrative fragments, explicitly or implicitly.

5.2.4 Virality by Character Type and by Number of Character-Roles

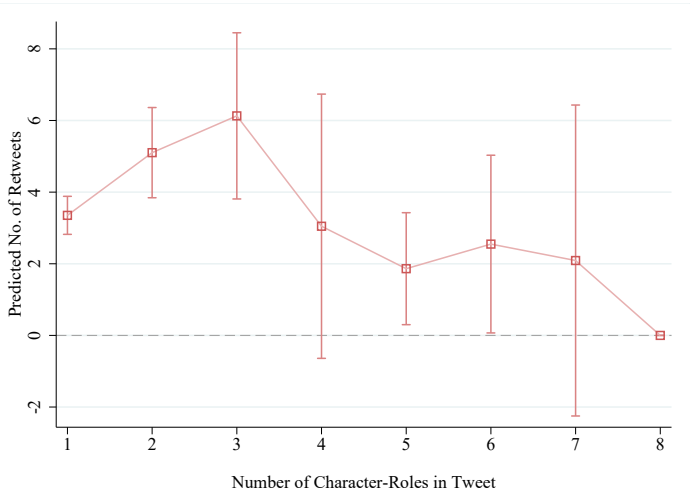
We finalize our analysis of political narrative virality by examining in more detail two other key options when drafting a narrative: What type of characters to feature and how many character-roles? Figure 5 showed individual characters descriptively, but here we more systematically analyze whether human or instrument characters tend to go more viral. Figure 9a, columns (1) and (2) compare tweets with only neutral characters with those with at least one human or instrument character in a drama triangle role. Column 2 adds the interaction between the two binary variables to test whether a combination of both is associated with a further boost to virality.

Figure 9: ***Regression Results – Heterogeneous Impact of Narratives on Virality***

(a) Heterogeneous Effects

Dependent Variable	**Retweets Count**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Human	0.652	0.553		
	(0.136)	(0.184)		
	[0.000]	[0.003]		
Instrument	0.245	0.094		
	(0.236)	(0.206)		
	[0.299]	[0.647]		
Human × Instrument		0.207		
		(0.299)		
		[0.489]		
Sum of CRs			0.214	0.659
			(0.057)	(0.150)
			[0.000]	[0.000]
Sum of CRs Squared				-0.125
				(0.045)
				[0.006]
Mean Outcome Reference Group	2.153	2.153	2.153	2.153
Pseudo R2	0.63	0.63	0.62	0.62
Obs.	309531	309531	309531	309531

(b) Marginal Effect of Adding Character-Roles



Notes: The exhibit provides an overview of the heterogeneous effects of narratives on virality. Figure 9a reports estimates from Poisson Pseudo-Maximum Likelihood regression models where the dependent variable is virality, measured by the number of retweets a tweet receives. Columns (1)–(2) isolate heterogeneity by character type. Column (1) includes dummy variables for tweets featuring at least one human and instrument character, with the omitted category being tweets containing only neutral characters. Column (2) augments this specification by including an interaction term identifying tweets that feature both human and instrument characters. Columns (3)–(4) explore the cumulative effect of featuring additional character-roles in a tweet- All regressions control for author characteristics (verified status, number of followers/followings, total tweets created, party affiliation, religiosity, higher education, and parenthood) and include character fixed effects. We also include hour, week-of-year, and year–state fixed effects. Standard errors are clustered at the week level, covering the full time frame. Table D.2 displays the results using an OLS model instead of Poisson Pseudo-Maximum Likelihood regression model. Figure 9b complements the regression results by plotting the marginal change in expected retweets associated with adding each additional character-role, providing an intuitive visualization of how increasing narrative complexity affects virality. Because of technical reasons with the Poisson model, this regression excludes the year-state FEs.

We find that human characters are associated with a systematically higher virality than instrument characters. The point estimate of having at least one human character-role is positive and clearly statistically significant, corresponding to a percentage change of 91.9%. The point estimate for instrument is positive, but only about a third the size and not significant at conventional levels. It would correspond to a percentage change of 27.8%. The interaction coefficient in column 2 is positive, indicating that a combination of human and instrument characters is fostering virality, but also far from being statistically significant. Considering only the magnitude, a combination of human with instrument yields the strongest effect on virality in percentage terms (113.8%).

These results can also be interpreted against the observed descriptive tendency of human characters becoming more frequent over the sample period relative to instrument characters. This could be regarded as a crowding out of discussions about potential solutions (policies, technologies, mechanisms) by more partisan discussions assigning credit and blame to the respective in- and out-group. For space reasons, we cannot go more in-depth here, but future research should examine these shifts and their most likely reasons.

Next, we examine the relationship between adding more character roles and virality. Column 3 uses the number of CRs as a regressor to test for the linear relationship, whereas column 4 also adds a quadratic term to examine potential nonlinearities. We do find that tweets with more CRs are on average significantly more viral, conditional on the set of controls and fixed effects in the main specification. However, column 4 indicates a non-linear relationship: at first adding more CRs boosts virality, but the marginal effect diminishes as more character-roles are added.

Figure 9b complements column 4 by plotting the predicted effects associated with additional character-roles. It turns out that adding up to 3 CRs boosts virality, but going to four and beyond leads to a clear decrease and partly insignificant point estimate. Combined with the prior results on role combinations, this effect has to be driven my multiple villains, combinations of heroes and villains and of villains and victims. At least suggestively, these patterns tentatively suggest certain strategies to boost virality: a stronger reliance on human characters, on villain characters, potentially combined with a hero or victim, but overall not more than three CRs.

6 Experimental Results: Beliefs, Preferences and Memory

While repeated and increased exposure through higher virality alone makes political narratives attractive as a communication technology, their effectiveness could be further enhanced by their persuasive power and a potential effect on memory. Using a large-scale observational dataset of tweets, we find that political narratives on average go more viral, with villain characters being even more efficient than hero characters. The well-documented mere-exposure effect in psychology shows that repeated exposure to the same message can already greatly enhance its effectiveness. However, is there also already an effect on beliefs, preferences and memory from single exposure to political narratives? To test this, we run three parallel survey experiments, each testing political narratives with different character-role combinations that were also frequent and viral on social media.

We distinguish between beliefs, preferences and memory effects. Narratives may be effective by shaping beliefs and thus influencing expectations and future decisions. Regarding preferences, we distinguish preferences about support for specific policies from preferences about characters. Policy preferences are generally more stable, and potentially tied to a partisan identity, and thus harder to be influenced by a single exposure to a specific narrative, while preferences about a character are more directly tied to the portrayal of that character in a drama triangle role in the political narrative. Changes in beliefs and in preferences about characters could be viewed as pre-requisites for changing concrete policy preferences.

The remainder of the section is organized as follows. First we provide an overview on the design of our experiments in Subsection 6.1, second we show our results on beliefs and preferences in Subsection 6.2, and thirdly, we explore the impact of political narratives on memory in Subsection 6.3.

6.1 Experimental Design

We conduct three pre-registered online experiments, each following an identical design. Our approach is in each case to test the effect of being exposed to a political narrative tweet compared to a control tweet that contains the same factual information and characters, but without portraying the characters in a drama triangle role. In the first experiment, the narrative tweet casts Green Tech and the US People as heroes. In the second, Green Tech again appears as a hero, joined by Regulations as a second hero, and Fossil Industry as a villain. In the third, Fossil Industry and Corporations are framed as villains, with Green Tech as the hero. Given the structure of the narratives, from here on we refer to the first experiment as the Hero-Hero (HH) experiment, the second as Hero-Hero-Villain (HHV) experiment, and the third as Villain-Villain-Hero (VVH) experiment.

In each experiment, participants are randomly assigned to either a treatment group – exposed to the narrative tweet – or a control group – exposed to the neutral version. We inform participants that the experiment investigates the effects of viewing a typical social media feed. All participants view a feed designed to resemble a Twitter/X timeline, with usernames blurred for anonymity. Each feed contains three posts. The first two – identical across both conditions – serve as obfuscation. The third post presents either the narrative tweet (in the treatment condition) or the neutral tweet (in the control condition). We recruit a representative sample of the US population via *Prolific*, randomly assigning participants to either the treatment or the control condition. Figure E.1 shows balance tests confirming no worrying differences between the treatment and control groups. To test memory effect, each experiment includes a follow-up survey administered one day later, with an overall attrition rate of roughly 22% (20% for the Hero-Hero experiment, 18% for the Hero-Hero-Villain, and 28% for the Villain-Villain-Hero). In this follow-up, all participants in both groups are shown the same feed as on the day before, but with the third (treatment or control) post blurred[11].

We use a combination of different types of questions, explained in more detail in the respective following sections. Our belief question asks participants to predict the share of renewable energy in the US by 2035. As this is unknown, there is no incentive to correctly guess; however, we also see no reason why the answers should be systematically biased. In the questions about policy preferences, the specific policy is adjusted to the specific treatment-control tweet combination. Given that changing preferences about the characters contained in narratives is a key channel of narrative persuasiveness that we highlight, we use an incentivized, real-stakes donation question for its measurement. For memory, we use a closed-form item-specific aided recall question with memory cues about numerical facts and an open, free recall question to measure character and role memory.

We begin by verifying that the treatment manipulation was effective. To do so, we use the answers to the open, free recall question on the day of the experiment, *"Please tell us anything you remember about the social media post that served as the basis of the previous questions. Describe your thoughts in the order they come to your mind; this can include full sentences, individual words, or attributes."* If the treatment manipulation was effective, treated participants should be more likely

[11] See the questionnaires of each experiment in Appendix E

Table 4: ***Manipulation Check - Effectiveness of the Political Narrative Treatment***

Dependent Variable	**Mention of Character-Roles**							
	Coeff./SE/p-value							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.293 (0.032) [0.000]	0.281 (0.032) [0.000]	0.254 (0.031) [0.000]	0.271 (0.018) [0.000]	0.358 (0.034) [0.000]	0.322 (0.035) [0.000]	0.280 (0.032) [0.000]	0.316 (0.019) [0.000]
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Outcome | Recall at least 1 Character					✓	✓	✓	✓
Experiment: Hero-Hero	✓			✓	✓			✓
Experiment: Hero-Hero-Villain		✓		✓		✓		✓
Experiment: Villain-Villain-Hero			✓	✓			✓	✓
Experiment FE				✓				✓
Mean Outcome Control Group	0.33	0.34	0.44	0.37	0.45	0.52	0.64	0.54
Observations	987	976	968	2,931	742	686	695	2,123

Notes: The table reports OLS estimates from manipulation checks of the narrative treatment. The dependent variable is a binary indicator equal to one if the participant recalled the role assigned to a character in the treatment tweet (e.g., Green Tech as hero in the Hero–Hero experiment, Fossil Industry as villain in the Hero–Hero–Villain experiment). Columns 1–4 report effects on the unconditional likelihood of recalling the character-role, while Columns 5–8 restrict the sample to participants who recalled at least one character from the tweet. Only the treatment group was exposed to characters explicitly framed in roles; control participants saw the same characters presented neutrally. Non-zero recall in the control group therefore reflects participants' prior beliefs or participants' interpretation of the control tweet, while the treatment effect captures the additional role attribution induced by framing. All regressions control for income, education, political preference, age, and sex, with standard errors clustered at the individual level. Appendix Table F.3 and Table F.9 report the corresponding models without controls and with randomization inference for p-values.

to mention the characters in the roles in which they were cast. Note than non-zero role assignment in the control group may reflect either prior beliefs or participants' subjective interpretation of the neutral character representation in the control tweet.

Table G.16 shows results separately for each experiment, Hero-Hero, Hero-Hero-Villain, and Villain-Villain-Hero, as well as for a pooled version with experiment fixed effects (FEs). The dependent variable is always 1 if the participant recalls at least one character-role correctly, and 0 otherwise. In columns (1) to (4), we show linear OLS models using this as an outcome without further restrictions. However, this bunches participants who remember the characters together with those that might not even remember the character. To examine specifically role perceptions more clearly, we thus also run specification in columns (5) to (8) that consider only participants who recall at least one character. The results are highly consistent across specifications, and slightly larger when conditioning on character recall. For instance, column 1 shows that compared to 33% in the control group, around 33+29=62% in the treatment group indicate at least one character in the correct role. The treatment manipulation work and those exposed to the political narrative version are approximately 30% more likely to mention the characters in a role.

6.2 Experimental Results: Beliefs and Preferences

We turn to the effects of the three political narratives on beliefs and stated preferences. Figure 10a shows the impact of political narratives on beliefs, separately for each treatment-control

pair: Hero–Hero (top panel), Hero–Hero–Villain (middle), and Villain–Villain–Hero (bottom). We show the coefficients for the belief question ("what percentage of US energy will come from green technologies in 2035") and the confidence in that estimate (ranging from 0 = not confident to 100 = very confident).

We find that the effects of the political narratives on beliefs are mirroring the roles and role combinations in the respective treatment. The Hero–Hero treatment with Green Tech and US People increases beliefs by about 2%, with the effect being statistically significant at the 10% level. However, combining Green Tech and Regulations as heroes with Fossil Industry in a villain role changes this positive outlook, which results in a near-zero and statistically insignificant effect. We then further emphasize the villain roles by casting two characters, Fossil Industry and Corporations, against Green Tech as hero. This completely turns around the effect on beliefs, with a negative point estimate of about 5% that is also statistically significant at the 1% level.

How should we interpret those effects on beliefs? First, political narratives clearly can shift beliefs even with single exposure. Second, the plausible pattern of belief changes across role combinations demonstrates the value of our framework for better understanding narratives and their effect. Thirdly, the effect of using villains, especially a villain combination, is by far the strongest, in line with our results on virality. Fourth, the only combination that increased confidence in the participants estimates is the fully aligned, non-antagonistic hero-hero role combination, while confidence for the antagonistic combinations remains unaffected. Overall, we conclude that single exposure to different narrative combinations does affect beliefs linked to a character-role conditional on the combination of that character with other character roles.

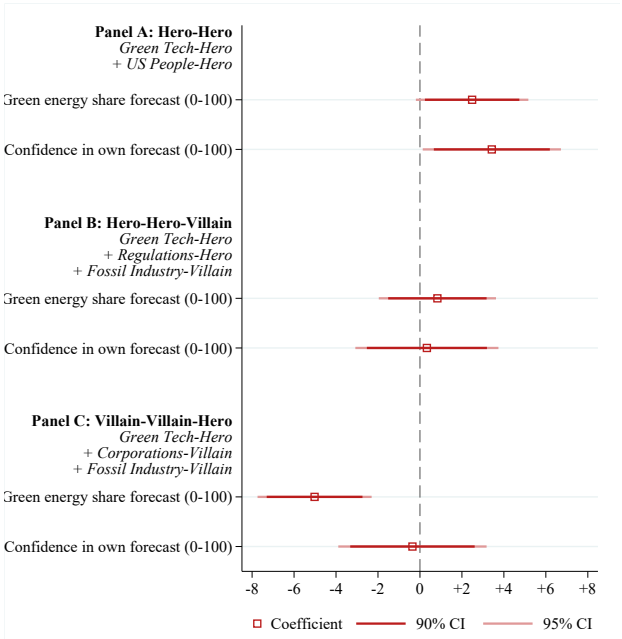
Moving to the right part of the figure, Figure 10b shows effects on stated policy preferences, using the same vertical structure as for beliefs. To do so, we measure support for specific climate change policy proposals that fit the respective character combinations in the treatment-control pairs. For the Hero–Hero combination, the question is about support for higher government subsidies for residential renewable energy systems. For the Hero–Hero–Villain combination, the policy proposal is about increasing the accountability of energy companies for potential damages they cause. For the Villain–Villain–Hero experiment, the question refers to support for raising taxes on fossil energy sources. We find no significant changes in policy support for any of the treatment-control combinations. This finding is consistent with existing research suggesting that policy preferences are less likely to be influenced by experimental interventions than beliefs (Berkebile-Weinberg et al. 2024).[12]

Finally, we examine a crucial potential lever of political narratives. Characters, from individual politicians to institutions, policies and technologies, are at the core of our approach of defining and measuring narratives. Hence, to understand the functioning of political narratives as a communication technology, and their persuasive power, it is essential to analyze whether (single) exposure can change preferences about the characters featured in the narrative? To test this, we drafted our

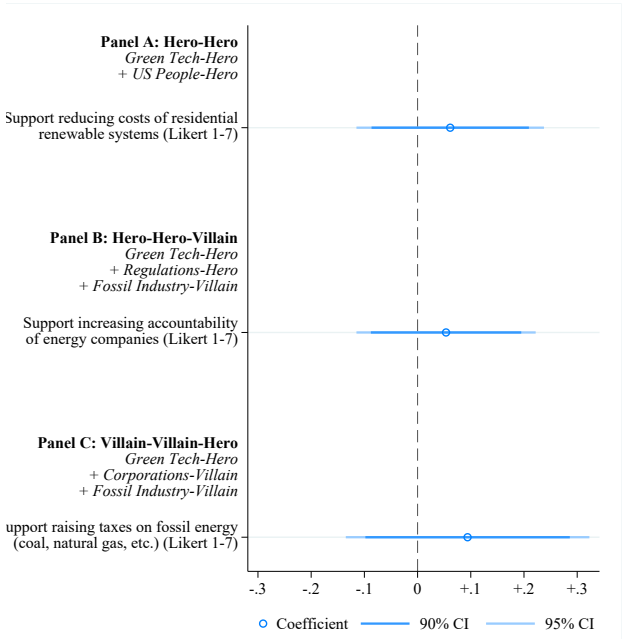
[12] In Table G.15, we also correlate narrative frequency with two policy preference question from the Cooperative Election Study (CES). We find that the frequency of Green Tech-Villain characters correlate negatively with expressed preferences in favor of renewable energy, and that the frequency of Regulation-Villain negatively with preferences in favor of the Environmental Protection agency. In contrast, the respective hero shares have an association close to zero and far from conventional levels of statistical significance.

Figure 10: ***Experimental Results - Impact of Political Narratives on Beliefs and Policy Preferences***

(a) Impact on Beliefs: Forecast for the Future



(b) Impact on Stated Policy Preferences: Policy Support

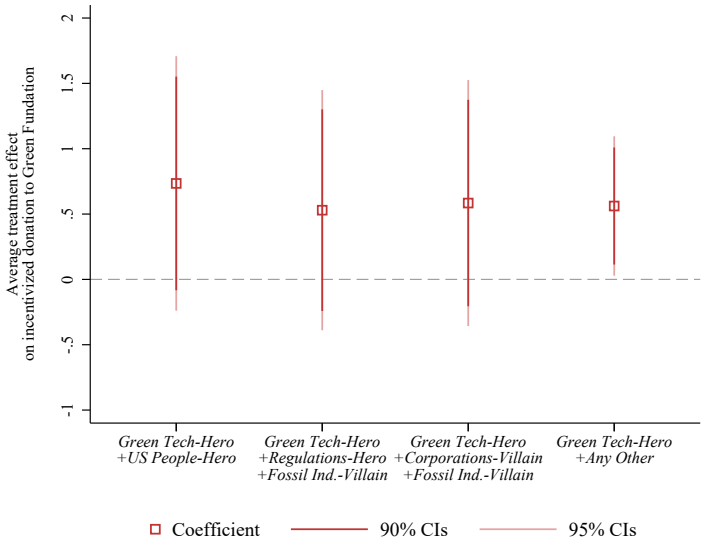


Notes: The figures show the coefficients from OLS regression models analyzing the impact of the narrative treatment on beliefs and stated preferences. In both figures, Panel A shows results for experiment Hero-Hero, Panel B for Hero-Hero-Villain and Panel C for Villain-Villain-Hero. In Figure 10a, we show the impact of the narrative treatment on two outcomes: the participants' forecast, as an answer to the question *'What percentage of US energy do you predict will come from renewable sources and green technology by the year 2035? Indicate a number between 0 and 100.'*, and their confidence in the forecast, as answer to *'Your response on the previous screen suggests that by 2035, [x]% of US energy will come from renewable sources and green technology. How certain are you that the actual share of renewable energy in 2035 will be between [x-5] and [x+5]%?'*. Table E.1, Table F.4, and Table F.10 show the corresponding regression models, with controls, without controls, and using randomization inference for p-values, respectively. In Figure 10b, we show the impact of the narrative treatment on the support or opposition for a policy or law that is in line with the content of the narrative. Policies questions are indicated in the graph and answers were collected as a 7-points Likert scale from 'Strongly Oppose' to 'Strongly Support'. All models include income, education, political preference, age, and sex as controls. We use robust standard errors. Table E.2, Table F.5, and Table F.11 show the corresponding regression models, with controls, without controls, and using randomization inference for p-values, respectively.

treatment-control pairs so that each contains Green Tech as a character, always in the hero role in the treatment condition. This enables us to consistently measure revealed preferences using a real-stakes, incentivized donation to a non-partisan organization supporting green technology as an outcome. Specifically, participants play a version of a dictator game in which they are asked to allocate $25 between themselves and the green technology organization. Donation outcomes like this tend to be statistically rather noisy relative to the expected effect size as many other factors influences the individual decision, hence it is important that our design allows us to also pool all three experiments to increase statistical power.

Figure 11 displays the coefficients together with 90% and 95% confidence intervals from each individual experiment, along with the pooled estimate for all experiments. he estimates consistently

Figure 11: ***Experimental Results - Impact of Political Narratives on Revealed Preferences about Character GREEN TECH***



Notes: The figure displays the coefficients, 90% (dark red), and 95% confidence intervals (light red) from OLS models analyzing the impact of the political narratives on participants' revealed preferences. We measure revealed preferences with an incentivized decision to donate 25$ to themselves or to a foundation promoting green technology diffusion. Moving from the left of the graph, coefficients 1, 2, and 3 show results for each experiment (in order Hero-Hero, Hero-Hero-Villain, and Villain-Villain-Hero), while the last coefficient on the right shows results for the pooled sample with experiment fixed effects. All models include income, education, political preference, age, and sex as controls. We use robust standard errors. Appendix Table E.3, Table F.6, and Table F.12 show the full corresponding regression models, with controls, without controls, and also when using randomization inference for p-values.

point in the same direction across experiments, suggesting a robust treatment effect of political narratives despite limited power in individual studies. Statistical power is limited in each individual experiment, because of substantial idiosyncratic noise in the donation outcome. However, exposure to the political narrative, each with Green Tech as the hero, does always increase participants' willingness to donate to the institution promoting green technology. Pooling across experiments, we find a positive effect that is statistically significant at the 5% level. On average, exposure to the Green Tech character in a hero role increases donations by around 0.56 dollars, with the overall average donation in the treatment condition being 7.02 dollars compared to 6.55 in the control condition. Accordingly, even a single exposure to a political narrative casting a characters in a specific drama triangle role can be sufficient to change a real-stakes decision related to that character.

6.3 Experimental Results: Memory

While our previous results indicate that a single exposure to a political narrative can shift beliefs and revealed preferences about a character, political narratives may also have an edge over other options to convey information with regard to information processing, storage, and retrieval. Graeber, Roth, and Zimmermann (2024) show that linking quantitative with qualitative information can improve

recall, especially if the memory retrieval question provides cues to the qualitative information. We investigate a related, but different question. We test (i.) whether the recall of numerical facts is improved when the facts are embedded in a political narrative, as well as (ii.) whether the recall of the characters is improved when they are framed in one of the drama triangle roles.

A single exposure to a political narrative may not be enough to shift policy preferences on its own. However, narratives can influence preferences incrementally, as repeated exposure combines with belief and attitude shifts triggered by even one exposure. To assess the potential for such cumulative effects, it is crucial to examine what kind of information from a single exposure is stored in memory and how well it can be recalled. Do political narratives mainly serve as a vehicle to communicate and anchor specific factual information, such as numerical data? Or is their primary strength in shaping perceptions of characters and their roles, thereby shifting preferences and beliefs more indirectly?

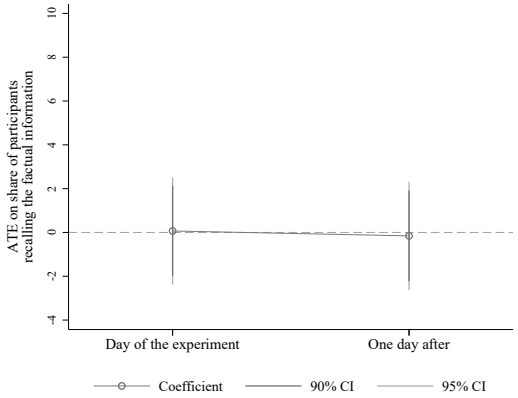
To investigate this, we included a memory recall task in the three experiments discussed above. We ask each respondent two information recall questions: first, a direct question with memory cues about the numerical fact contained in the texts, and second, a free-recall question about anything else they remember about the text they were exposed to. Both questions were posed both on the day of the experiment and again in a follow-up survey conducted a day later[13].

Figure 12a and Figure 12b show the results of the exposure to political narratives on memory, pooling all three treatment-control pairs. The left panel shows the coefficient measuring the effect of being exposed to the narrative treatment on remembering the factual information contained in the tweet, the right panel the effect on remembering characters contained in the tweets. The panels show that political narratives positively affect memory recall of at least one character, with the effect also being clearly statistically significant. Memory decay between the two days is minimal (from around 6% to around 5%), and the effect remains significant. In contrast, political narratives have no effect on factual memory, with the effect on recall of numerical information being very close to zero and clearly insignificant. Those null effects are robust to using different target ranges instead of exact numerical recall as the outcome (see Appendix Table F.1 and Table F.2). Hence, it seems the main strength of political narratives as a communication technology with regard to memory is that the narrative characters, in their respective roles, are much better remembered by recipients.

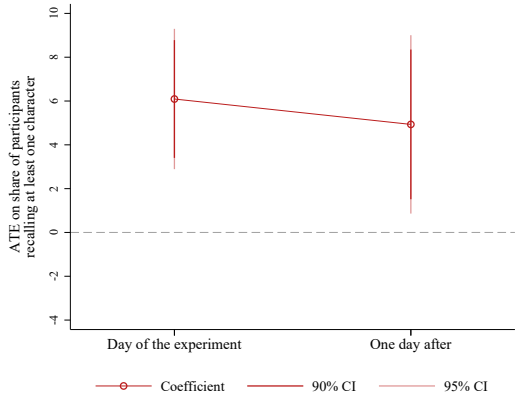
[13]The question posed on the day of the experiment serves as basis for the results of our manipulation check as well, in Table G.16.

Figure 12: ***Experimental Results - The Impact of Political Narratives on Memory (Pooled Sample)***

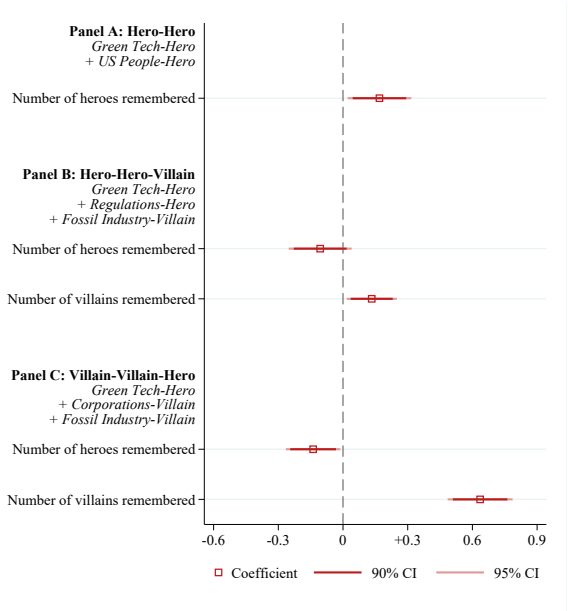
(a) Factual Memory:
Recall of Numerical Information



(b) Character Memory:
Recall of at least one character



(c) Character Memory:
Recall of characters by role



Notes: The figure reports OLS estimates of the treatment effect on participants' memory. Figure 12a shows the pooled impact on factual recall from the tweet on the day of the experiment (left) and one day later (right), expressed as percentage differences between treatment and control. Figure 12b shows the effect on recalling at least one character from the text, again same day (left) and next day (right), using an open-ended recall question coded into a binary outcome. Appendix Table E.4, Table F.7, and Table F.13 show the corresponding regression models, with controls, without controls, and using randomization inference for p-values, respectively. Figure 12c examines recall of specific characters: Panel A (Hero–Hero) tests recall of Green Tech and US People (heroes); Panel B (Hero–Hero–Villain) shows recall of Green Tech and Regulations (heroes, top) and Fossil Industry (villain, bottom); Panel C (Villain–Villain–Hero) shows recall of Green Tech (hero) and Fossil Industry/Corporations (villains). All models control for income, education, political orientation, age, and sex; SEs are clustered at the participant level. The first two figures also include experiment fixed effects. Appendix tables report the corresponding models with controls, without controls, and using randomization inference for p-values. Appendix Table E.5, Table F.8, and Table F.14 show the corresponding regression models, with controls, without controls, and using randomization inference for p-values, respectively.

Our framework allows us to go beyond this general memory effect and distinguish between the effect on characters cast as heroes compared to those cast as villains in the different treatment-control combinations. Figure 12c differentiates for each treatment-control combination between the effect on remembering those characters portrayed as hero from those portrayed as villain in the treatment condition. The first panel shows that solely using two characters in hero roles enhances memory of both characters, as expected. However, the second and third panel reveal an interesting relationship between hero and villain role memory. Once a character cast as villain is inserted, memory of the hero characters is crowded out and actually lower than in the control tweets. In contrast, recall of the villain characters is always enhanced, even more strongly if two characters are cast into the villain role.[14] Hence, casting characters as villains not only enhances virality and persuasiveness, but in addition even improves memory recall of that character at the expense of characters in other roles.

Discussion and Interpretation

This section discusses possible explanations specifically for the memory results, with some explanations also possibly relevant for the prior experimental results. We do not go into more detail here as this would go beyond the scope of the paper, but we are convinced that there is ample room for more research to investigate the channels and mechanisms in more detail.

a.) Information resonance. People are more likely to remember information that resonates with them and with their whole set of memories. Many of these memories are in the form of stories and narratives, and naturally feature characters in the three drama triangle roles. Hence, the representation of characters in these roles plausibly has an inherent advantage for encoding information and storing it efficiently into memory. We can think of the roles in the treatment as adding meaning to the characters, which improve the encoding and retrieval of the characters in memory. The schema theory in psychology also highlights how using default, existing schemas to simplify complex information can be seen as a mechanism for efficient information storage and easier retrieval. Participants in the treatment condition also remember the roles (as we show in Table G.16), in line with the idea that they store the character and its role jointly in their memory.

b.) Cue-similarity. A key insight in Graeber, Roth, and Zimmermann (2024) is that cue similarity between a story and the prompt/question use for memory retrieval plays a key role in further enhancing the better memory of stories. Our differential results between fact and character recall do not seem to be driven by cue similarity. In fact, our question about the fact ("How many billion kWhs were generated using solar energy in 2023 in the US? Please indicate your best guess.") has many more cues to the fact that our open ended question from which we infer characters ("Please tell us anything you remember about the social media post"). Hence, it is noteworthy that we find the improved character recall even with such a free-recall question.

c.) Type of recall. Our fact question is a verbatim numeric recall of a specific number,

[14] The character outcome is coded from open-ended free recall. We cannot exclude that item-specific cued recall of heroes would attenuate the crowding-out pattern. Accordingly, we put more emphasis on the difference between the effect on characters in hero versus villain roles than on the absolute effect size.

which are generally harder to remember for most people. This number in our context is linked to a character (Green Tech), but it is linked to that character in both treatment and control. While our manipulation adds qualitative role content to the text, similar to Graeber, Roth, and Zimmermann (2024), this role assignment is not directly number-focused. Thus, our results show that such verbatim recall of a specific number is not further improved by assigning a role to that character. The better memory of characters and roles can actually be regarded as in line with Graeber, Roth, and Zimmermann (2024), who also document better recall of information type and direction, not specific numbers.

d.) Attention. Attention influences what information people process, but also what and how they store it (Loewenstein and Wojtowicz 2025). Attention is drawn towards aspects of a text that are salient. The drama triangle roles can be stimulating in that sense for a person because they reflect a salient category, e.g. a threat (Vuilleumier 2005). Similarly, when "searching" their memory for information retrieval, prior evidence indicates attention guides people how to search for information. If a character in memory is linked to a larger semantic cluster, like an archetypal role, that could help to retrieve it more easily.

e.) Emotions. Although it is plausible that emotions play an important role also for memory formation, a closer look at our results reveal that the character-role combinations forming the political narratives capture more than just emotion. We can distinguish valence (how positive relative to negative a text is) from specific types of emotions like joy, fear or anger (as used in e.g., Algan et al. 2025). Figure E.2 shows the valence difference of each treatment vs. control tweet comparison, using the same standard dictionaries as before, as well as the coefficient from the experimental results discussed above. Valence is in line with the changes in beliefs, with more negative treatments triggering a decline in beliefs. However, the donation outcomes related to Green Tech is consistently positive and almost identical in size across the three treatment-control pairs, demonstrating that it is not the valence of the whole tweet that matters, but the role assignment of the respective character. There is also no direct relationship between the level of anger and the experimental results on belief and preferences. The memory results further demonstrate that it is the specific character-role combinations that drive the effects, not simply valence or emotions.

f.) Strategic implication for political communication. Our experiments show that single-shot narrative exposure reliably imprints who the actors are and how they are cast, but does not improve recall of facts, even when those facts are tied to a key character. This possibly gives some leeway to the sender of such information in terms of spreading imprecise or even false facts. If most recipients of political narratives only remember the characters and their roles better, but not the facts, this lowers the costs of spreading false facts. Ex post fact-checking only partly solves this problem, as the original narrative - thanks to its virality - is usually read much more widely than possible corrections. New approaches like community notes are more interesting in that regard, as they tie the fact-checking directly to the political narrative. At least to the extent the narratives remains with the social media platform, is shared and displayed together with it.

7 Conclusion

In conclusion, our study demonstrates that the political narrative framework provides a powerful lens for understanding how narratives drive engagement and influence public opinion. The result is a numerical map of the narrative citizens share, one that economists can merge with behavioral and market data. By observing which characters and roles dominate, we gain predictive leverage over both the direction and the intensity of public debate.

By analyzing US climate change policy discussions on Twitter over a decade, we show what makes narratives go viral. Political narratives, on average, are about sixty per cent more likely to be retweeted. This result holds controlling for a range of time and region fixed effects, for key authors characteristics like the number of followers, and when using character fixed effects. Negativity or emotions alone cannot explain that virality premium: when we hold eight discrete emotions and continuous valence constant, the effect on virality is only slightly decreased and remains clearly significant. A villain framing lifts retweets by roughly 170 per cent, hero framing by about 55 per cent, while victim framing has little effect. Pairing another role with a villain raises virality, whereas adding other roles has an ambiguous effect, indicating that complexity can impose an attention cost. Human characters generally tend to go more viral than instrument characters like technology or policies.

Across three preregistered surveys we embed single narrative tweets in an otherwise ordinary feed and compare the effect to a tweet with the same characters in neutral roles. A one-time exposure to the political narrative shifts beliefs: respondents adjust their expectations about the character in the direction implied by the narrative. It also nudges revealed preferences: when given an incentive-compatible choice, participants reallocate real money toward or away from the actor highlighted in the story, even though their stated policy support remains effectively unchanged from single exposure. One day later participants reliably remember which characters appeared more often, yet are not more likely to reproduce factual numbers that accompanied the text, showing that political narratives are more about characters than facts.

Economists value causal narratives for showing how one action leads to another, capturing an important aspect of narratives as a communication technology. Causal narratives trace sequences of events – taxes raise prices, prices curb emissions, monetary expansion causes inflation. Political narratives add an explicit assignment of agency, blame, and moral standing to relevant human (politicians, institutions) or instrument characters (technologies, policies). Corporations become villains, households victims, activists heroes, and the very same chain of events acquires a specific purpose and direction. Because these role labels can shift while the underlying causality stays fixed, political narratives add an orthogonal dimension that pure event-based stories cannot capture. They also reach far beyond raw emotion: a sentence may sound negative without naming a culprit or emphasize a hero without using strong emotion. Our evidence shows that the presence of a villain, not the amount of negativity, is what multiplies viral reach and shapes expectation. Distinguishing characters and roles from both causality and sentiment can therefore be essential for understanding

how ideas travel and influence decision-making.

As political narratives increasingly shape public discourse and attention, understanding their structure, spread, and impact will be essential for economists, policymakers, and platforms alike. Our framework applied with the suggested pipeline outputs structured data with explicit character and role variables, enabling researchers to study narratives in any large text data set and easily combine it with other economic or political data. Because characters and the archetypal drama triangle roles are so fundamental to human story-telling, standard LLMs are really efficient in measuring well-defined character-roles, making this also a very cost-effective way of measuring narratives without manual coding. Possible applications beyond our context are widespread. For instance, macroeconomists could now test whether villain framing of central banks widens inflation-expectation tails; finance scholars can observe how firms move from hero to villain status during scandals and how this affects stock market evaluations; development economists can monitor shifts in donor narratives about recipient governments and the role of aid agencies. Measurement, once the bottleneck, no longer stands in the way of such inquiries.

References

Acemoglu, Daron, Tarek A Hassan, and Ahmed Tahoun (2018). "The Power of the Street: Evidence from Egypt's Arab Spring". In: *The Review of Financial Studies* 31.1, pp. 1–42.

Adena, Maja et al. (2021). "Bombs, Broadcasts and Resistance: Allied Intervention and Domestic Opposition to the Nazi Regime During World War II". In: *SSRN Electronic Journal*.

Akerlof, George A. and Dennis J. Snower (2016). "Bread and Bullets". In: *Journal of Economic Behavior & Organization* 126.Part B, pp. 58–71.

Algan, Yann et al. (2025). "Emotions and Policy Views". In: *Working Paper*.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005). "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools". In: *Journal of Political Economy* 113.1. Publisher: The University of Chicago Press, pp. 151–184.

Andre, Peter et al. (2024). "Misperceived Social Norms and Willingness to Act Against Climate Change". In: *Review of Economics and Statistics*, pp. 1–46.

Andre, Peter et al. (2025). "Narratives about the Macroeconomy". In: *The Review of Economic Studies*.

Anker, Elisabeth (2005). "Villains, Victims and Heroes: Melodrama, Media, and September 11". In: *Journal of Communication* 55.1, pp. 22–37.

Aridor, Guy et al. (2024). "The Economics of Social Media". In: *Journal of Economic Literature* 62.4, pp. 1422–1474.

Ash, Elliott and Sergio Galletta (Oct. 2023). "How Cable News Reshaped Local Government". In: *American Economic Journal: Applied Economics* 15.4, pp. 292–320.

Ash, Elliott, Germain Gauthier, and Philine Widmer (2024). "RELATIO : Text Semantics Capture Political and Economic Narratives". In: *Political Analysis* 32.1, pp. 115–132.

Ash, Elliott and Mikhail Poyker (2024). "Conservative News Media and Criminal Justice: Evidence from Exposure to Fox News Channel". In: *The Economic Journal* 134.660, pp. 1331–1355.

Ash, Elliott et al. (2021). "Visual Representation and Stereotypes in News Media". In: *Center for Law & Economics Working Paper Series* 2021.15.

Ash, Elliott et al. (2024a). "From Viewers to Voters: Tracing Fox News' Impact on American Democracy". In: *Journal of Public Economics* 240, p. 105256.

Ash, Elliott et al. (2024b). "The Effect of Fox News on Health Behavior during COVID-19". In: *Political Analysis* 32.2, pp. 275–284.

Barrera, Oscar et al. (2020). "Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics". In: *Journal of Public Economics* 182, p. 104123.

Barron, Kai and Tilman Fries (2024). "Narrative Persuasion". In: *WZB Discussion Paper No. SP II 2023-301r*.

Baylis, Patrick (2020). "Temperature and Temperament: Evidence from Twitter". In: *Journal of Public Economics* 184, p. 104161.

Beach, Brian and W Walker Hanlon (2023). "Historical Newspaper Data: A Researcher's Guide and Toolkit". In: *Explorations of Economic History* 90, p. 101541.

Berger, Jonah (2011). "Arousal Increases Social Transmission of Information". In: *Psychological Science* 22.7, pp. 891–893.

— (2016). *Contagious: Why Things Catch On*. Simon and Schuster.

Berger, Jonah and Katherine L Milkman (2012). "What Makes Online Content Viral?" In: *Journal of Marketing Research* 49.2, pp. 192–205.

Bergstrand, Kelly and James M Jasper (2018). "Villains, Victims, and Heroes in Character Theory and Affect Control Theory". In: *Social Psychology Quarterly* 81.3, pp. 228–247.

Berkebile-Weinberg, Michael et al. (2024). "The Differential Impact of Climate Interventions Along the Political Divide in 60 Countries". In: *Nature Communications* 15.1, p. 3885.

Braghieri, Luca et al. (2024). "Article-Level Slant and Polarization of News Consumption on Social Media". In: *SSRN 4932600*.

Bursztyn, Leonardo et al. (2023). "Opinions as Facts". In: *The Review of Economic Studies* 90.4, pp. 1832–1864.

Bénabou, Roland, Armin Falk, and Jean Tirole (2020). "Narratives, Imperatives, and Moral Reasoning". In: *NBER Working Paper No. 24798*.

Caesmann, Marcel et al. (2021). "Going Viral: Propaganda, Persuasion and Polarization in 1932 Hamburg". In: *Discussion Paper DP16356*.

Caesmann, Marcel et al. (2024). *Censorship in Democracy*. Version Number: 1.

Cage, Julia et al. (2022). "Hosting Media Bias: Evidence from the Universe of French Broadcasts, 2002-2020". In: *Working Paper*.

Cagé, Julia (2020). "Media Competition, Information Provision and Political Participation: Evidence From French Local Newspapers and Elections, 1944–2014". In: *Journal of Public Economics* 185, p. 104077.

Cagé, Julia, Nicolas Hervé, and Marie-Luce Viaud (2020). "The Production of Information in an Online World". In: *The Review of Economic Studies* 87.5, pp. 2126–2164.

Cagé, Julia et al. (July 2023). "Heroes and Villains: The Effects of Heroism on Autocratic Values and Nazi Collaboration in France". In: *American Economic Review* 113.7, pp. 1888–1932.

Chen, Jiafeng and Jonathan Roth (2024). "Logs with Zeros? Some Problems and Solutions". In: *The Quarterly Journal of Economics* 139.2, pp. 891–936.

Chu, Zi et al. (2012). "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" In: *IEEE Transactions on dependable and secure computing* 9.6, pp. 811–824.

Dechezleprêtre, Antoine et al. (Apr. 2025). "Fighting Climate Change: International Attitudes toward Climate Policies". In: *American Economic Review* 115.4, pp. 1258–1300.

Djourelova, Milena and Ruben Durante (2022). "Media Attention and Strategic Timing in Politics: Evidence from U.S. Presidential Executive Orders". In: *American Journal of Political Science* 66.4, pp. 813–834.

Djourelova, Milena, Ruben Durante, and Gregory J Martin (2025). "The Impact of Online Competition on Local Newspapers: Evidence from the Introduction of Craigslist". In: *Review of Economic Studies* 92.3, pp. 1738–1772.

Djourelova, Milena et al. (2024). "Experience, Narratives, and Climate Change Beliefs". In: *SSRN Electronic Journal*.

Durante, Ruben and Brian Knight (2012). "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy". In: *Journal of the European Economic Association* 10.3, pp. 451–481.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei (2019). "The Political Legacy of Entertainment TV". In: *American Economic Review* 109.7, pp. 2497–2530.

Durante, Ruben and Ekaterina Zhuravskaya (2018). "Attack When the World Is Not Watching? US News and the Israeli-Palestinian Conflict". In: *Journal of Political Economy* 126.3, pp. 1085–1133.

Eliaz, Kfir and Ran Spiegler (2020). "A Model of Competing Narratives". In: *American Economic Review* 110.12, pp. 3786–3816.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova (2020). "Social Media and Protest Participation: Evidence from Russia". In: *Econometrica* 88.4, pp. 1479–1514.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya (2011). "Media and Political Persuasion: Evidence from Russia". In: *American Economic Review* 101.7, pp. 3253–85.

Esposito, Elena et al. (June 2023). "Reconciliation Narratives: *The Birth of a Nation* after the US Civil War". In: *American Economic Review* 113.6, pp. 1461–1504.

Fog, Klaus et al. (2010). "The Four Elements of Storytelling". In: *Storytelling*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 31–46.

Gehring, Kai, Joop Age Adema, and Panu Poutvaara (2022). "Immigrant Narratives". In: *CESifo Working Paper No. 10026*.

Gentzkow, Matthew and Jesse M Shapiro (2010). "What Drives Media Slant? Evidence from US Daily Newspapers". In: *Econometrica* 78.1, pp. 35–71.

Gomez-Zara, Diego, Miriam Boon, and Larry Birnbaum (2018). "Who Is the Hero, the Villain, and the Victim? Detection of Roles in News Articles Using Natural Language Techniques". In: *23rd International Conference on Intelligent User Interfaces*, pp. 311–315.

Graeber, Thomas, Christopher Roth, and Florian Zimmermann (2024). "Stories, Statistics, and Memory". In: *The Quarterly Journal of Economics* 139.4, pp. 2181–2225.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart (Mar. 2023). "Designing Information Provision Experiments". In: *Journal of Economic Literature* 61.1, pp. 3–40.

Halberstam, Yosh and Brian Knight (2016). "Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter". In: *Journal of Public Economics* 143, pp. 73–88.

Harari, Yuval Noah (2014). *Sapiens: A Brief History of Humankind*. Random House.

Jiangli, Su (2020). "European & American Think Tanks and the Reality of US-China Trade War: An NPF Application". In: *Journal of Social and Political Sciences* 3.2.

Jones, Michael D (2014). "Cultural Characters and Climate Change: How Heroes Shape our Perception of Climate Science". In: *Social Science Quarterly* 95.1, pp. 1–39.

Jones, Michael D. and Mark K. McBeth (2010). "A Narrative Policy Framework: Clear Enough to Be Wrong?" In: *Policy Studies Journal* 38.2, pp. 329–353.

Karpman, Stephen (1968). "Fairy Tales and Script Drama Analysis". In: *Transactional Analysis Bulletin* 7.26, pp. 39–43.

Kendall, Chad and Constantin Charles (2022). "Causal Narratives". In: *NBER Workin Paper 30346*.

Kirilenko, Andrei P and Svetlana O Stepchenkova (2014). "Public Microblogging on Climate Change: One Year of Twitter Worldwide". In: *Global Environmental Change* 26, pp. 171–182.

Lagakos, David, Stelios Michalopoulos, and Hans-Joachim Voth (2025). "American Life Histories". In: *NBER Working Paper 33373*.

Landis, J. Richard and Gary G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data". In: *Biometrics* 33.1, p. 159.

Loewenstein, George and Zachary Wojtowicz (2025). "The Economics of Attention". In: *Journal of Economic Literature* 63.3, pp. 1038–1089.

Macaulay, Alistair and Wenting Song (2022). "Narrative-Driven Fluctuations in Sentiment: Evidence Linking Traditional and Social Media". In: *No. 2023-23. Ottawa: Bank of Canada*.

Merry, Melissa K (2016). "Constructing Policy Narratives in 140 Characters or Less: The Case of Gun Policy Organizations". In: *Policy Studies Journal* 44.4, pp. 373–395.

Michalopoulos, Stelios and Christopher Rauh (2024). "Movies". In: *NBER Working Paper 32220*.

Michalopoulos, Stelios and Melanie Meng Xue (2021). "Folklore". In: *The Quarterly Journal of Economics* 136.4, pp. 1993–2046.

Mohammad, Saif M. and Peter D Turney (2013). *NRC Emotion Lexicon*. Tech. rep. Artwork Size: 234 p. National Research Council of Canada, 234 p.

Montoya, R. Matthew et al. (2017). "A Re-Examination of the Mere Exposure Effect: The Influence of Repeated Exposure on Recognition, Familiarity, and Liking". In: *Psychological Bulletin* 143.5, pp. 459–498.

Müller, Karsten and Carlo Schwarz (2021). "Fanning the Flames of Hate: Social Media and Hate Crime". In: *Journal of the European Economic Association* 19.4, pp. 2131–2167.

— (2023). "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment". In: *American Economic Journal: Applied Economics* 3, pp. 270–312.

Oehl, Bianca, Lena Maria Schaffer, and Thomas Bernauer (2017). "How to Measure Public Demand for Policies When There is no Appropriate Survey Data?" In: *Journal of Public Policy* 37.2, pp. 173–204.

O'Brien, Erin (2018). *Challenging the Human Trafficking Narrative: Victims, Villains and Heroes*. Routledge.

Polletta, Francesca et al. (2011). "The Sociology of Storytelling". In: *Annual Review of Sociology* 37.1, pp. 109–130.

Qian, Nancy and David Yanagizawa-Drott (2017). "Government Distortion in Independently Owned Media: Evidence from U.S. News Coverage of Human Rights". In: *Journal of the European Economic Association* 15.2, pp. 463–499.

Serra-Garcia, Marta (2025). "The Attention–Information Tradeoff". In: *CESifo Working Paper No. 11885*.

Shiller, Robert J. (Apr. 2017). "Narrative Economics". In: *American Economic Review* 107.4, pp. 967–1004.

— (2020). *Narrative Economics: How Stories Go Viral and Drive Major Economic Events*. Princeton University Press.

Tabassum, Fatima et al. (2023). "How Many Features Do We Need to Identify Bots on Twitter?" In: *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity*. Ed. by Isaac Sserwanga et al. Cham: Springer Nature Switzerland, pp. 312–327.

Terry, Larry D (1997). "Public Administration and the Theater Metaphor: The Public Administrator as Villain, Hero, and Innocent Victim". In: *Public Administration Review*, pp. 53–61.

Voth, Hans-Joachim and David Yanagizawa-Drott (2025). "Image(s)". In: *CEPR Discussion Paper* DP19219.

Vuilleumier, Patrik (2005). "How Brains Beware: Neural Mechanisms of Emotional Attention". In: *Trends in Cognitive Sciences* 9.12, pp. 585–594.

Yanagizawa-Drott, David (2014). "Propaganda and Conflict: Evidence from the Rwandan Genocide". In: *The Quarterly Journal of Economics* 129.4, pp. 1947–1994.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020). "Political Effects of the Internet and Social Media". In: *Annual Review of Economics* 12, pp. 415–438.

Virality:
What Makes Narratives Go Viral, and Does it Matter?

Kai Gehring* Matteo Grigoletto**

Appendix

*University of Bern, Wyss Academy at the University of Bern, e-mail: *ka.gehring@unibe.ch*
**University of Bern, Wyss Academy at the University of Bern, e-mail: *matteo.grigoletto@unibe.ch*

Appendix

Table of Contents

A Process and methods

This appendix provides additional details on our pipeline. The aim is to facilitate its replicability and assist researchers in using our methodology for similar projects. While some details may overlap with those mentioned in the paper, this appendix provides complementary information.

A.1 Data extraction

This section outlines the data source and time frame of the extracted data. It is important to recognize potential differences in narrative structures in various sources, such as books, newspapers, and social networks. While digitized newspapers (Gehring, Adema, and Poutvaara 2022; Beach and Hanlon 2023) and social media (Cagé, Hervé, and Viaud 2020) are common sources of text data in economics, other formats, like transcribed TV, radio, YouTube broadcasts, or open-ended survey responses, also offer valuable material. Our framework is adaptable to any type of text.

In this study, we focus on English-language tweets from the United States, posted on Twitter (now X) between 2010 and 2021. At the time this project began, the historical Twitter APIv2 allowed researchers access to all tweets posted (and not deleted) since 2006. We chose the US because Twitter plays a significant role in policy discussions, and 2010 marks the point when Twitter became a mainstream platform. While the sample of US Twitter users is not fully representative, it offers a unique opportunity to observe the creation and spread of narratives over time and across different regions.

Keywords and query

We indicate here the keywords and rules used for the query of Twitter historical APIv2. Consider the following conditions:

1. The tweet includes at least one of the following terms: 'climate change', 'global warming', 'renewable energy', 'energy policy', 'emission', 'certificate trading', 'green certificate', 'white certificate', 'combined heat', 'power solution', 'energy solution', 'CO2', 'energy efficiency', 'energy saving', 'solar power', 'solar energy', 'wind power', 'wind energy', '*renewable energies*', '*energy policies*', '*ipcc*', '*green growth*', '*green-growth*', '*green wash*', '*green-wash*', '*climate strike*', '*climate action*', '*strike 4 climate*', '*strike for climate*'.
2. The tweet includes at least one of the following terms: 'climate', 'global warming', 'greenhouse' AND at least one of the following terms: 'refining', 'feed-in', 'cogeneration', 'extraction', 'exploitation', 'geotherm', 'hydro', 'agriculture', 'waste management', 'forest', 'wood', '*problem*', '*issue*', '*effect*', '*gas*', '*degrowth*', '*de-growth*', '*fridaysforfuture*', '*fridays4future*', '*scientistsforfuture*', '*scientists4future*'.
3. The tweet includes at least one of the following terms: 'climatechange', 'globalwarming', 'renewableenergy', 'renewableenergies', 'energypolicy', 'energypolicies', 'greencertificate', 'whitecertificate', 'combinedheat', 'powersolution', 'energysolution', 'energyefficiency', 'energysaving', 'solarpower', 'solarenergy', 'windpower', 'windenergy', 'greengrowth', 'greenwash', 'climatestrike', 'climateaction', 'strike4climate', 'strikeforclimate'.

4. The tweet includes term 'carbon' AND Tweet does NOT include any of the following terms: 'bicycle', 'bike', 'copy', 'fiber', 'rims', 'altered', 'fork', 'frame', 'dating', 'tacos'.

A tweet is part of our sample if any of the above conditions applies. In addition, a tweet is part of our sample if its text also satisfies all of the following:

- (a) The tweet does not contain an URL address.
- (b) The tweet's content is in English language.
- (c) The tweet is not a retweet.

The following are the changes adopted in deviation from keywords and rules proposed by Oehl, Schaffer, and Bernauer (2017), the paper of reference for us to define our query:

1. In Oehl, Schaffer, and Bernauer (2017) any keyword needs to appear in combination with at least one among: 'climate', 'global warming', 'greenhouse'. We do not adopt the condition as baseline but we use it for those words that refer to climate change in a looser way (see condition No. 2 above).
2. We use some terms that are not present in Oehl, Schaffer, and Bernauer (2017): 'ipcc', 'climate change', 'energy policies', 'renewable energies', 'green growth', 'green-growth', 'green wash', 'green-wash', 'climate strike', 'climate action', 'strike 4 climate', 'strike for climate', 'problem', 'issue', 'effect', 'gas', 'degrowth', 'de-growth', 'fridaysforfuture', 'fridays4future', 'scientistsforfuture', 'scientists4future' (see words in *italics* in conditions 1 and 2).
3. We use all multi-word expressions in condition 1 (e.g. 'energy policies') also as hashtags (see condition 3).
4. We use an exclusion restriction tailored towards tweets, because we realized there was a consistent pattern of false positive cases with the word 'carbon' (see condition 4).

Extracted data

In this analysis, we use data extracted via the Historical Twitter API in two distinct ways. First, we collected tweets from randomly selected days within the time frame of interest, which we refer to as the 'random days' dataset. Second, we gathered tweets from every Saturday within the same period, which we call the 'every Saturday' dataset. These two datasets were then combined into a final dataset. The random days dataset aims to provide a representative sample of tweets across the entire period. The inclusion of tweets from every Saturday helps to capture data that is less likely to be influenced by specific events, unless those events are cyclical and consistently occur on Saturdays.

Random days dataset

We collect tweets extracted from a set of randomly selected days in the period 2010-2021. We use the calendar option of the online random number generator random.org to randomly select a day

within each month of this time period. The extraction was done on 7^{th} and 8^{th} February 2022. The selected days are the following:

2010-01-28, 2010-02-14, 2010-03-13, 2010-04-30, 2010-05-25, 2010-06-30, 2010-07-07, 2010-08-04, 2010-09-14, 2010-10-02, 2010-11-06, 2010-12-14, 2011-01-14, 2011-02-09, 2011-03-01, 2011-04-10, 2011-05-13, 2011-06-19, 2011-07-22, 2011-08-15, 2011-09-08, 2011-10-23, 2011-11-21, 2011-12-17, 2012-01-31, 2012-02-27, 2012-03-26, 2012-04-04, 2012-05-26, 2012-06-18, 2012-07-10, 2012-08-18, 2012-09-20, 2012-10-22, 2012-11-01, 2012-12-03, 2013-01-15, 2013-02-12, 2013-03-27, 2013-04-25, 2013-05-05, 2013-06-18, 2013-07-19, 2013-08-08, 2013-09-25, 2013-10-11, 2013-11-06, 2013-12-01, 2014-01-24, 2014-02-13, 2014-03-04, 2014-04-30, 2014-05-16, 2014-06-23, 2014-07-12, 2014-08-21, 2014-09-26, 2014-10-24, 2014-11-05, 2014-12-06, 2015-01-26, 2015-02-21, 2015-03-20, 2015-04-24, 2015-05-06, 2015-06-09, 2015-07-23, 2015-08-20, 2015-09-15, 2015-10-15, 2015-11-11, 2015-12-21, 2016-01-11, 2016-02-05, 2016-03-22, 2016-04-02, 2016-05-01, 2016-06-19, 2016-07-01, 2016-08-31, 2016-09-09, 2016-10-13, 2016-11-14, 2016-12-22, 2017-01-01, 2017-02-12, 2017-03-25, 2017-04-04, 2017-05-07, 2017-06-05, 2017-07-11, 2017-08-27, 2017-09-14, 2017-10-21, 2017-11-09, 2017-12-21, 2018-01-09, 2018-02-09, 2018-03-30, 2018-04-06, 2018-05-08, 2018-06-05, 2018-07-06, 2018-08-14, 2018-09-16, 2018-10-22, 2018-11-12, 2018-12-15, 2019-01-04, 2019-02-14, 2019-03-15, 2019-04-19, 2019-05-17, 2019-06-21, 2019-07-22, 2019-08-30, 2019-09-19, 2019-10-01, 2019-11-01, 2019-12-01, 2020-01-12, 2020-02-12, 2020-03-22, 2020-04-16, 2020-05-08, 2020-06-22, 2020-07-17, 2020-08-17, 2020-09-26, 2020-10-08, 2020-11-07, 2020-12-18, 2021-01-23, 2021-02-25, 2021-03-20, 2021-04-05, 2021-05-23, 2021-06-12, 2021-07-11, 2021-08-30, 2021-09-25, 2021-10-10, 2021-11-11, 2021-12-30.

Every Saturday dataset

We collect a large sample of tweets extracted over the same period of analysis 2010-2021. We collect tweets from every Saturday of every week between January 2010 and December 2021. The extraction was done between 4^{th} and 7^{th} December 2022.

Data managing

We compute a number of steps to clean and organize data after the extraction. We describe these steps in the following points:

1. Despite setting API's filter, some non-English tweets were captured and we had to clean them using langdetect, a python port of the language-detection library in Java. At the time of writing, langdetect is also available as an extension in spaCy.

2. The text of a single tweet might satisfy more than one condition of our query, hence representing a potential duplicate in the extracted data. Each tweet is associated to a uniquely identifying ID that we use to drop potential duplicates.

3. Before labeling, we clean the tweets of emojis and any other unicode objects that have a UTF-8 code larger than three bits. We also replace line-breaks in the text with simple spaces.

The random days dataset - after the cleaning and wrangling - comprises 1,070,702 tweets. The every Saturday dataset - after the cleaning and managing - consists of 3,279,730.

A.2 Data geo-localization

The tweets in our datasets were posted by users from around the world. Since our interest in this paper focuses on discussions in the United States, we filter the tweets to include only those originating from the US before proceeding with the annotation. In the following, we provide some indications on localization of tweets.

It is important to notice that tweets do not inherently come with localization information and this needs to be retrieved by the researchers, if possible. Many authors using tweets in their analysis developed their own methods to localize tweets (Kirilenko and Stepchenkova 2014; Baylis 2020). We build on previous work and structure our own method that exploits different 'fields' of information provided by the APIv2.

There are two main sources of geographical information available through the Historical API. Among the available user fields, there is one called 'location'. This can be filled in two ways. One method is directly by the user who decides to indicate her location when creating the profile. Another is by the Twitter API algorithm itself, which detects the location if it has been mentioned in the text of the user's self-description. For example, if a user describes herself as 'I am a PhD student based in Zurich', the API would provide Zurich as the user's location. Additionally, among the available tweet fields, there is one called 'geo'. This indicates the location of the tweet if the tweet has been geo-tagged somewhere. In fact, the Twitter application allows users to tag a tweet with a specific location at the time of posting. This simply involves indicating a place to which the user wants to 'tag' the post.

We decide to prioritize the information about the user and hence assign to each tweet the location of the user that posted it. This is because only a minority of tweets come with 'geotag' information. This might raise the suspect that people tag their tweets only during particular events - such as holidays or work trips - which do not truly represent their environment/location. Only in cases where the user's location information is unavailable do we locate a tweet according to the geo-tag assigned to it, if any. Consequently, our localization pipeline consists of the following steps, which apply to the analysis dataset:

1. We collect all available locations relative to users' profiles from the two datasets 'random days' and 'every Saturday'.

2. We use the geopy implementation of Nominatim's API which exploits OpenStreetMap data. We query the API inputting the location in string format and obtain its geographical coordinates.

3. Once each string location is associated to a set of coordinates we merge the locations back to the datasets. The merge is done with the formula 'many to one' so that users with the same location are associated to the same set of coordinates.

4. We repeat step 1, 2 and 3 for those tweets that could not be located by the description location of the users posting them but that present a geo-tag.

5. For all tweets that could be located (either via description or via geo-tag location) we intersect their coordinates with a shapefile of the United States borders and keep only those tweets located within the country.

6. The two concatenated dataset count a total of 4,236,799 tweets, after dropping potential duplicates. Once filtered for location, keeping only tweets originating from the US, the total amount is 1,151,693 tweets.

Some important notes on the Nominatim API. First, the API algorithm returns the centroid coordinates of the location, hence when searching for e.g. 'Florida' it would return the coordinates of the centroid of the state of Florida. Second, when the string location is not clear the API returns 'NaN' output. Third, in most of the cases in which the location string is composed of two or more locations - e.g. 'Florida and NY' - the API returns either one of the two or 'NaN' output. This is generally hard to predict, but multiple locations are a minority of the cases in our data. Lastly, the API is not case-sensitive hence e.g. 'New York City' and 'new york city' would provide the same coordinates.

A.3 OpenAI API annotation

This section provides a guideline for the process used to annotate data via the OpenAI API. The following steps summarize the procedure with key details about the data involved in each phase.

Setting up the OpenAI API

The first step is to create a project on the OpenAI API platform. Ideally, this should be done using a business account to ensure maximum data privacy. Upon project creation, an API Key is generated, which is required for all queries and operations using the OpenAI API. The API Key can be found in the user's profile under the API Key Dashboard.

Data organization

The annotation pipeline takes as input the set of data that was labeled as originating from the US: a total of 1,151,693 tweets.

Annotation modality

For each input tweet we query the OpenAI API, prompting the selected model (more about this below) to annotated the tweet according to our instructions. The annotation happens in two separate stages. In the first stage, we prompt the selected model to categorize the tweet's relevance to the climate change policy discussion in the US. In particular we distinguish the following labels:

- `irrelevant`: When the tweet is not really about climate change. E.g. '*The political climate is getting very day more heated!!*'
- `assert`: When the tweet is about climate change but it is limited to asserting the existence of the issue, without touching onto any adaptation or response policy or action. E.g. '*#Climatechange is the single most important issue we are facing, wake up!*'

- `deny`: When the tweet is about climate change but it is limited to denying its existence or proposing sarcastic and/or skeptical view on the extent of the problem. E.g. '*Where is this ""global warming"" when one needs it?? It's cold outside, climate is NOT changing.*'
- `policy`: When the tweet is about climate change and it discusses related policies and issues. E.g. '*Not recognizing climate change is an issue is just insane, we need to start supporting policies that actually make a change like the carbon tax.*'

Stage 2 consists in the individuation of characters and the classification of their role, only for those tweets that were labeled as `policy` in stage 1. In particular, we query the model to find the following characters:

- Developing Countries: emerging and developing economies, poorer countries, or nations part of the BRICS group -Brazil, Russia, India, China, South Africa-, as well as any related government institutions, representatives, or citizens associated with these countries and regions.
- US Democrats: politicians, members, and public figures associated with the US Democratic Party. This includes prominent individuals such as Joe Biden, Nancy Pelosi, Alexandria Ocasio-Cortez, Bernie Sanders, Barack Obama, and others who represent ideals and policies of the Democratic party.
- US Republicans: politicians, members, and public figures associated with the US Republican Party. This includes prominent individuals such as Trump, Mitch McConnell, Ted Cruz, Ron DeSantis, and others who represent ideals and policies of the Republican Party.
- Corporations: large corporations, small and medium businesses, banks, and other private sector entities. This includes CEOs and leadership figures like Elon Musk and Jeff Bezos, representatives from industries such as energy, technology, finance, and manufacturing, as well as industry lobbying groups, corporate interests, and small or local business owners
- US People: the collective citizens, voters, workers, youth, and grassroots movements of the United States, often portrayed in contrast to political elites or corporate interests. This also includes references to the 'average American,' and general terms like the public, society, community action, and any collective expression of public will or activism, including movements like FridaysForFuture, Sunrise Movement, and Extinction Rebellion
- Emission Pricing: any market-based instruments and schemes designed to price carbon emissions and incentivize reductions. This includes tools such as carbon taxes, cap and trade systems, carbon pricing, emission trading, carbon markets, pollution credits, carbon credits, carbon fees, and carbon dividends. These tools are part of the broader market-led response to addressing climate change.
- regulations: government actions, policies and movements aimed at combating climate change through the banning, phasing out, or strict regulation of specific products or industries. This includes efforts such as banning fracking, phasing out fossil fuels, banning single-use plastics, and other regulatory measures designed to reduce environmental impact

and promote sustainability. It also encompasses movements that challenge economic growth models or advocate for systemic changes, such as de-growth movements and anti-capitalist environmental initiatives.

- FOSSIL INDUSTRY[15]: any explicit reference to fossil fuels, including terms like coal, oil, natural gas, and related critical labels such as 'dirty energy.' This encompasses all forms of energy derived from fossil sources, as well as the technologies and infrastructure that rely on them, such as combustion engines, power plants, and industrial machinery.
- GREEN TECH: technologies developed as a response to climate change or aimed at phasing out fossil fuels. This includes wind energy, solar energy, electric vehicles, hydrogen power, battery storage, geothermal energy, and other renewable or low-carbon technologies excluding though nuclear energy.
- NUCLEAR TECH: all forms of nuclear energy, including both fusion and fission technologies. This encompasses nuclear power plants, nuclear reactors, nuclear fusion research, and related technologies used for energy production.

For each character, the model determined if the character was present and whether they played the role of *hero*, *villain*, *victim*, or *neutral* (if no clear role applied). Those tweets that were classified as featuring at least a character among our listed characters, form the so-called `relevant` tweets dataset, the main dataset used in our analysis, comprising a total of 309,744 tweets.

For both stages, the prompts were generated using OpenAI's ChatGPT interface. One of the authors queried the GPT-4o model in ChatGPT, providing an explanation of the task and asking the model to suggest the optimal prompt for instructing itself. Since the GPT-4o model is the same used via the API, this method ensured efficient and effective prompt construction. The final prompts used for the annotation are provided below:

Stage 1 prompt:

You are an average US citizen. The user will provide the content of a tweet posted from the US. Your task is to analyze the tweet within the context of US political discourse, particularly in relation to climate change. Respond in JSON format. 1. Relevance Check: Analyze the tweet in the context of US climate change discussion and determine its relevance. Provide one of the following values: - 0 (irrelevant): If the tweet does not discuss climate change in a meaningful way. For example, if it only includes a hashtag (like #climatechange) or a passing reference but does not engage in any discussion about climate change or related policies, it should be considered irrelevant. - 1 (assert): If the tweet asserts the existence of climate change but does not engage with specific policies or actions related to it. This includes tweets that acknowledge climate change as an issue without going deeper into details. - 2 (deny): If the tweet denies the existence or severity of man-made climate change, referring to it as a hoax, scam, or fraud, or using sarcasm or language that undermines the reality of climate change. - 3 (relevant): If the tweet discusses climate change or related policies in a substantive way. This includes any tweet that debates,

[15]We refer to the character FOSSIL INDUSTRY in the paper as FOSSIL FUELS in the prompt.

critiques, or supports policies or actions related to climate change, as well as conversations on how to combat or adapt to climate change. Respond in JSON format, returning the value in the key 'r'.

Stage 2 prompt:

You are an average US citizen. The user will provide the content of a tweet posted from the US between 2010 and 2021. Your task is to analyze it within the context of US political discourse, particularly in relation to climate change and related policies. Respond in JSON format.1. Character Analysis: Identify whether the tweet mentions specific characters. For each character mentioned, assess their contextual role using the following scale: - Villain (1): The character is portrayed as contributing to problems, opposing positive change, negatively or engaging in harmful actions related to climate change. Look for language that blames, criticizes, or attributes a negative impact. - Hero (2): The character is portrayed as leading efforts to combat climate change, promoting environmental policies, positively or acting in a morally commendable way. Look for praise, leadership roles, or proactive efforts. - Victim (3): The character is portrayed as being unfairly attacked, facing challenges, or suffering due to external factors. Look for language that depicts them as unjustly targeted, enduring consequences, suffering or being the victim. - No role (4): Choose this option if the tweet mentions the character but does not clearly assign one of the above described roles, or if the context is ambiguous or neutral.2. Character Definitions: Evaluate these characters in the context of the tweet. - For Developing Countries and Emerging Economies (emerging and developing economies, poorer countries, or nations part of the BRICS group -Brazil, Russia, India, China, South Africa-, as well as any related government institutions, representatives, or citizens associated with these countries and regions), provide the assessment in the key 'a': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For The US Democrats (politicians, members, and public figures associated with the US Democratic Party. This includes prominent individuals such as Joe Biden, Nancy Pelosi, Alexandria Ocasio-Cortez, Bernie Sanders, Barack Obama, and others who represent ideals and policies of the Democratic party), provide the assessment in the key 'b': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For The US Republicans (politicians, members, and public figures associated with the US Republican Party. This includes prominent individuals such as Trump, Mitch McConnell, Ted Cruz, Ron DeSantis, and others who represent ideals and policies of the Republican Party), provide the assessment in the key 'c': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Corporations and Industry (large corporations, small and medium businesses, banks, and other private sector entities. This includes CEOs and leadership figures like Elon Musk and Jeff Bezos, representatives from industries such as energy, technology, finance, and manufacturing, as well as industry lobbying groups, corporate interests, and small or local business owners), provide the assessment in the key 'd': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For The People of the US (the collective citizens, voters, workers, youth, and grassroots movements of the United States, often portrayed in contrast to political elites or corporate interests. This also includes references to the 'average

American,' and general terms like the public, society, community action, and any collective expression of public will or activism, including movements like FridaysForFuture, Sunrise Movement, and Extinction Rebellion), provide the assessment in the key 'e': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Emission Pricing Tools (any market-based instruments and schemes designed to price carbon emissions and incentivize reductions. This includes tools such as carbon taxes, cap and trade systems, carbon pricing, emission trading, carbon markets, pollution credits, carbon credits, carbon fees, and carbon dividends. These tools are part of the broader market-led response to addressing climate change), provide the assessment in the key 'f': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Banning or Regulation Policies (government actions, policies and movements aimed at combating climate change through the banning, phasing out, or strict regulation of specific products or industries. This includes efforts such as banning fracking, phasing out fossil fuels, banning single-use plastics, and other regulatory measures designed to reduce environmental impact and promote sustainability. It also encompasses movements that challenge economic growth models or advocate for systemic changes, such as de-growth movements and anti-capitalist environmental initiatives), provide the assessment in the key 'g': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Fossil Fuels (any explicit reference to fossil fuels, including terms like coal, oil, natural gas, and related critical labels such as 'dirty energy.' This encompasses all forms of energy derived from fossil sources, as well as the technologies and infrastructure that rely on them, such as combustion engines, power plants, and industrial machinery.), provide the assessment in the key 'h': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Green Technologies (technologies developed as a response to climate change or aimed at phasing out fossil fuels. This includes wind energy, solar energy, electric vehicles, hydrogen power, battery storage, geothermal energy, and other renewable or low-carbon technologies excluding though nuclear energy), provide the assessment in the key 'i': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies. - For Nuclear Energy (all forms of nuclear energy, including both fusion and fission technologies. This encompasses nuclear power plants, nuclear reactors, nuclear fusion research, and related technologies used for energy production), provide the assessment in the key 'j': - 0: No mention of the character. - 1: Villain. - 2: Hero. - 3: Victim. - 4: None of the roles applies.3. Final Output: Respond with a JSON format containing the following keys: - 'a': (0-4 as defined in step 2). - 'b': (0-4 as defined in step 2). - 'c': (0-4 as defined in step 2). - 'd': (0-4 as defined in step 2). - 'e': (0-4 as defined in step 2). - 'f': (0-4 as defined in step 2). - 'g': (0-4 as defined in step 2). - 'h': (0-4 as defined in step 2). - 'i': (0-4 as defined in step 2). - 'j': (0-4 as defined in step 2).

OpenAI API Batch Modality

The final dataset used for annotation was large, and annotating it using the standard OpenAI API endpoints would have been too time-consuming. To speed up the process, we used the Batch API, which allows users to upload a large number of requests at once. OpenAI processes these requests

within 24 hours, optimizing for times of lower traffic.

The size of each batch depends on the prompt and input, and more details can be found on the Batch API page. In our case, we uploaded 25,000 tweets per batch, resulting in 61 chunks for the entire dataset. Users can monitor the status of each batch on their Dashboard. Our entire dataset was processed in about a week, with a total cost of approximately 2,100 USD.

Lastly, regarding data retention: OpenAI does not use uploaded data for model training and retains it only for legal reasons, currently for one month. We recommend deleting files created via the Batch API through the Dashboard after processing.

A.4 Requirements and Sources

Table A.1: *Data Sources*

Data	Source	Download Date	Availability
Twitter Data			
Model-tweets dataset	Twitter Historical APIv2	7^{th} - 8^{th} Jan. 2022	Cannot be shared
Analysis-tweets dataset	Twitter Historical APIv2	4^{th} - 7^{th} Dec. 2022	Cannot be shared
GIS Data			
US Shapefile (v4.1)	GADM website	4^{th} Oct. 2022	Can be shared
Survey Data			
Coperative Election Study (CES)	Tufts University	11^{th} Apr. 2025	Can be shared

Notes: The table reports a description of the sources of the data used in our analysis and their respective availability.

B Narrative Classification: GPT vs Human Coders

Artificial Intelligence technology has undertaken a revolution in recent years, influencing many human activities and tasks. Among these, research is widely changing, especially when the tasks and goals include the interpretation, generation, and understanding of human language. Models like GPT are increasingly more used to classify, summarize, and extract meaning from text. These tools offer researchers a fast and scalable way to analyze large volumes of language data.

In our study, we apply GPT to a complex task: understanding and classifying political narratives shared by users on Twitter. Despite the drama triangle (Karpman 1968) being a widely adopted model of storytelling, deeply rooted in human communication, narratives may still leave some space for interpretation. In comparison, other NLP tasks tend to rely more on a clear "ground truth". For example, a model may classify whether a review is positive or negative, or whether a message contains offensive language. In the case of narratives, even trained Human Coders can disagree on whether a particular text expresses a given narrative (see, for example, the exploration in Gehring, Adema, and Poutvaara (2022)).

This appendix compares GPT's classifications to those of Human Coders on the same tweets. We treat GPT's output as reflecting an 'average representative' Human Coder. Rather than a validation exercise, this comparison explores how closely GPT's interpretations align with those of human annotators. Below, we describe the method and present results at two levels: the character-role level and the tweet level.

Method

We start this comparison exercise by hiring workers from Amazon MTurk. To guarantee high-quality human coding, we design a qualification task to select attentive workers. The task requires participants to read the instructions for our study and answer four comprehension questions to assess their understanding. Only those who correctly answer all four questions are invited to proceed to the actual coding task. Additionally, we include a Captcha verification step to deter potential bots.

The aim of the exercise is the classification of 500 tweets, randomly selected among the tweets identified by GPT as containing at least one character (`relevant` tweets). We structured the MTurk assignment so that each tweet was classified by two Human Coders, allowing us to compute measures of inter-coder reliability among them. In total, 80 workers successfully completed the qualification test and were invited to participate. Of these, 28 workers actually classified tweets. Workers were free to decide how many tweets they would classify. Some coded as few as one tweet, while others classified up to 130. On average, workers classified 36 tweets each. We kept the task open until we reached the objective of each of the 500 tweets being coded by two different workers.

The classification task was divided into two phases, matching as closely as possible the language and structure of the prompt used for GPT's task. For each character, the Human Coder was first asked whether the character was present in the tweet. For characters identified as present, coders then indicated whether the character was depicted as a hero, villain, victim, or none of these roles.

Finally, we used these classifications to compare Human Coders to each other and to GPT. Below, we present the results.

Comparison at the Character Level

In the first part of this analysis, we compare classifications at the character level. This reflects the structure of the classification task, where Human Coders were asked, for each tweet, to evaluate each of the ten characters individually – first judging whether the character was present, then assigning a role if applicable. To mirror this structure, we organize the dataset so that each line corresponds to one character within a tweet. As a result, for each of the 500 tweets in the comparison exercise, the dataset contains ten lines, one for each character.

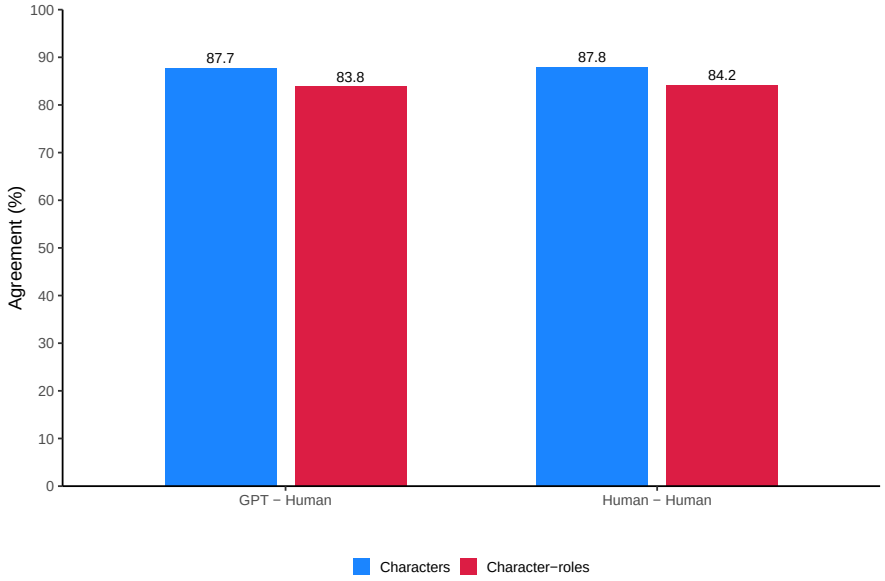
As a first step, we compare the overall agreement between GPT and Human Coders, and between pairs of Human Coders. Figure B.1 summarizes the results. The figure shows the share of tweets where GPT and the Human Coder agreed on the presence of the character (blue) and on the presence of the character-role (red), shown in the first two columns from the left. Importantly, GPT is not compared to the same Human Coder across all tweets but to whichever coder classified each specific tweet. The next two columns report the same agreement rates, but for pairs of Human Coders who coded the same tweet. In all comparisons, agreement includes negative agreement, thus cases where both coders (or GPT and a Human Coder) agreed that the character or character-role was absent.

Overall, we find a high level of agreement between Human Coders and GPT. On average, GPT and Human Coders classified the presence of characters the same way in 87.7% of cases. Agreement on character-role classifications was similarly high, at 83.8%. Notably, these rates closely mirror the agreement between Human Coders themselves. Two independent Human Coders agreed on the character in 87.8% of cases and on the character role in 84.2% of cases. These results support our view of GPT as an average or representative Human Coder.

As explained above, our main agreement measure includes both positive and negative classifications. In most cases, characters are **not** present, so including negative agreement (where coders agree a character is absent) can inflate the overall agreement rates. To ensure the results are not driven by these cases, we compute Fleiss' κ, which adjusts for agreement that may occur by chance. The Fleiss' κ amounts to 0.578 for agreement on character, and 0.486 for the agreement on character-roles. These correspond to a moderate level of agreement according to the conventional interpretation by Landis and Koch (1977). Even after correcting for chance agreement and the prevalence of negative classifications, the level of agreement remains relatively high, further supporting the reliability of both human and GPT-based coding.

In the second step of our analysis, we examine the agreement on the assignment of drama triangle roles (Karpman 1968). As we argue in the paper, the drama triangle is not just a communication tool deeply rooted in the history of storytelling, but also a natural way humans interpret reality. Based on this, we expect high levels of agreement when it comes to assigning roles. In the previous analysis (Figure B.1), agreement captured both uncertainty about the presence of characters and

Figure B.1: ***Overall Agreement on Characters and Character-Roles***



Notes: The figure shows agreement rates between GPT and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected tweets. All tweets were classified by GPT as containing at least one of characters of interest in our study. Twenty-eight Human Coders, each coding a different number of tweets, classified the sample. Each tweet was coded by two Human Coders. The first two bars (left) show the share of tweets where GPT and the Human Coder agreed on the presence of the character (blue) and the character-role (red). GPT is compared to whichever Human Coder classified each tweet. The next two bars show the same agreement rates between the two Human Coders who coded each tweet. Agreement includes both positive and negative cases, meaning instances where coders (or GPT and a Human Coder) agreed that a character or role was absent.

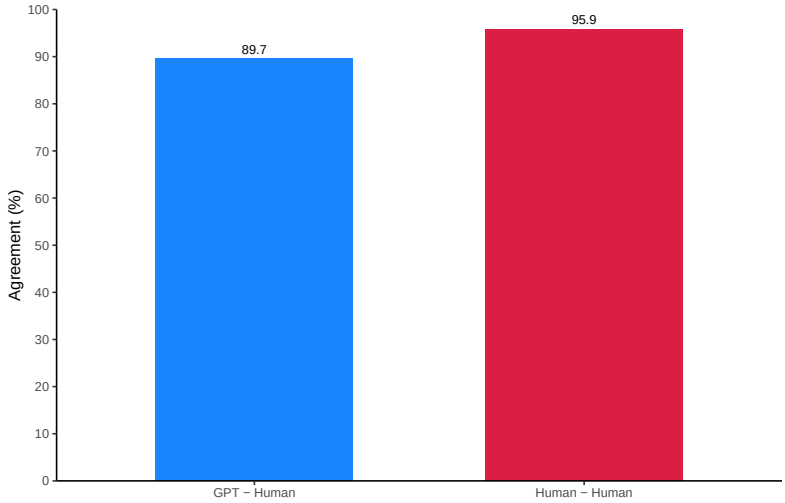
the assignment of roles. In other words, coders were compared on both whether a character was present and which role, if any, was assigned. In the next step, we select only those tweets where the two Human Coders agreed that a specific character was present, to then compare the agreement on the assignment of roles, both between the Human Coders and between GPT and Human Coders.

Figure B.2 shows the results of this second exercise. On the left, the blue bar indicates agreement between GPT and Human Coders; on the right, the red bar shows agreement between pairs of Human Coders, limited to tweets where both Human Coders agreed on the presence of the character. As expected, agreement levels are high: Human Coders agreed in nearly 96% of cases, while GPT agreed with Human Coders in almost 90% of cases. We compute also in this case the Fleiss' κ which measures 0.668 in this case, indicating very high agreement, as expected.

Comparison at the Tweet Level

In the second part of this analysis we compare classifications at the tweet level. This entails using the tweets classified by GPT and Human Coders, to build variables mirroring those used in the analysis, and then assess the level of agreement on these variables. Figure B.3 displays the results, on which we provide further details below.

Figure B.2: ***Agreement on Character-Role Conditional on Human Coders Agreeing on the Presence of Characters***



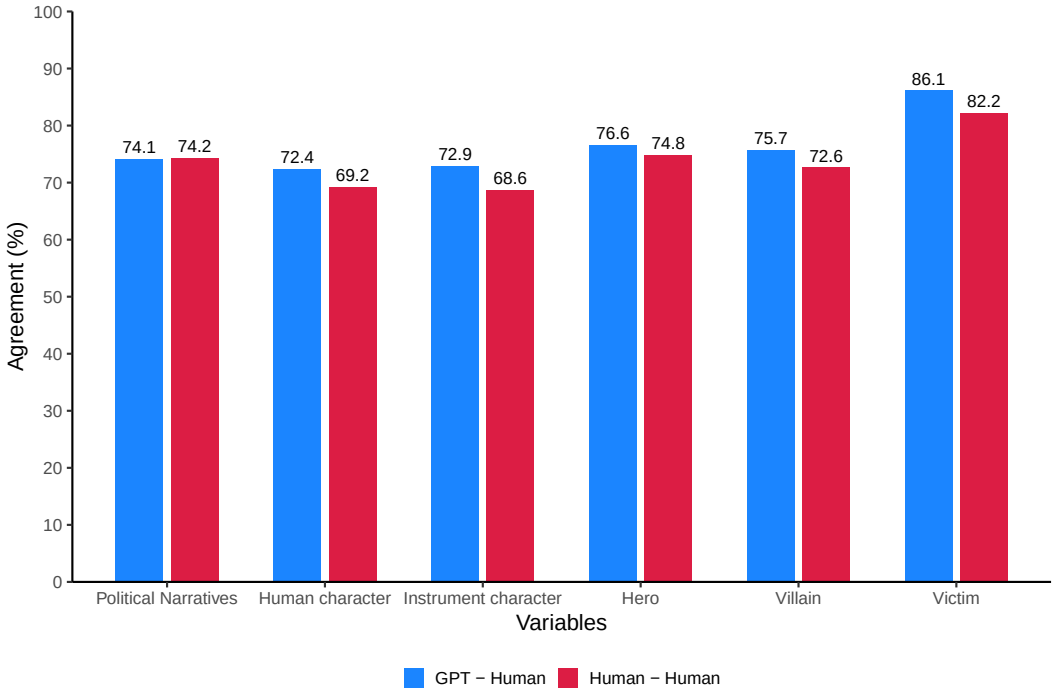
Notes: The figure shows agreement rates between GPT and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected tweets. All tweets were classified by GPT as containing at least one of characters of interest in our study. Twenty-eight Human Coders, each coding a different number of tweets, classified the sample. Each tweet was coded by two Human Coders. For each tweet the dataset comprises ten entries, one for each character of interest. For this exercise we retain those entries of the dataset where the two Human Coders agreed on the characters' presence. The left column shows the share of these tweets for which GPT and the Human Coders agreed on the assignment of the roles to characters. The right column shows the same for the agreement between Human Coders.

Political Narratives We compute the Political Narratives variable at the tweet level, defined as containing at least one character-role combination. We assess agreement rates between GPT and Human Coders (blue) and between pairs of Human Coders (red). As shown in Figure B.3, agreement levels are very similar: around 74% for GPT–Human Coder comparisons and 75% for Human Coder pairs. As above, we also compute Fleiss' κ to provide an unbiased measure. The κ value is 0.213, indicating fair agreement.

Human and Instrument Characters We use the classified tweets to compute two additional variables also used in our analysis: the Human Character variable, defined as containing at least one human character, and the Instrument Character variable, defined as containing at least one instrument character. We observe a high level of agreement in detecting both human and instrument characters. As shown in Figure B.3, on average, GPT and Human Coders agreed on the presence of at least one human character in approximately 73% of tweets. Similarly, the inter-human agreement on detecting both human and instrument characters is about 69%. Notably, two Human Coders agree, on average, to a lower degree than when pairing GPT with any Human Coder. Fleiss' κ in this case is higher, at 0.428, indicating moderate agreement.

Hero, Villain, and Victim Roles Finally, we explore agreement on variables capturing the classification of roles. We construct three variables: Hero, Villain, and Victim, defined respectively as containing at least one hero, villain, or victim character-role in the tweet. This analysis further

Figure B.3: ***Agreement on Measures Mirroring the Variables Used in the Analysis***



Notes: The figure shows agreement rates between GPT and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected tweets. All tweets were classified by GPT as containing at least one of characters of interest in our study. Twenty-eight Human Coders, each coding a different number of tweets, classified the sample. Each tweet was coded by two Human Coders. Agreement is defined as the number of identical classifications over the number of total tweets. *Political narratives* is defined as the level of agreement on the presence of at least one character-role in a tweet (Hero, Villain, Victim). *Human character* is defined as agreement on the presence of at least one human character. *Instrument character* is defined as agreement on the presence of at least one instrument character. *Hero*, *Villain*, *Victim* measures agreement for the presence of at least one character-role in the tweet. The blue bars indicate the average level of agreement between GPT and the two humans. The red bar indicates the average level of agreement between two humans.

supports the validity of our approach. Figure B.3 shows that GPT and Human Coders agreed on the presence of heroes in roughly 76% of cases, villains in 76%, and victims in about 86%. Once again, Human Coders agreed with GPT more often than they agreed with each other. Fleiss' κ values are 0.508 for heroes, 0.493 for villains, and 0.255 for victims. The lower κ for victims likely reflects the lower frequency of victim roles and the uneven distribution of this classification, which often takes a value of zero. This makes it more likely to agree 'by chance' on this particular role.

C Additional Output: Twitter sample

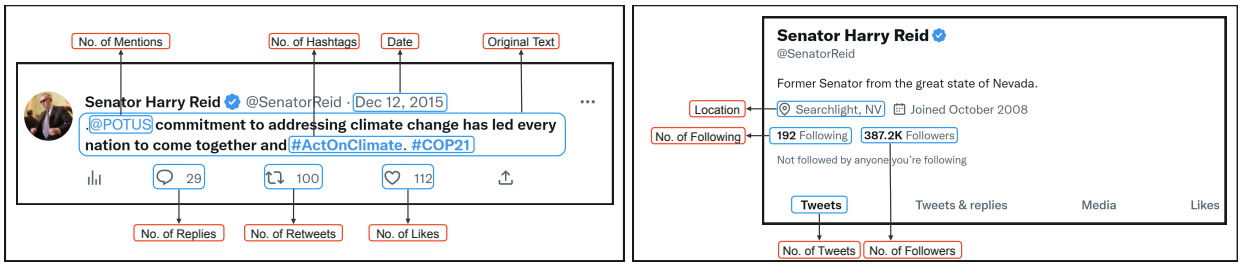
C.1 Additional Details on Variables

In this section of Appendix C, we provide additional information on the variables used in our analysis. We start with Figure C.1, that offers a visual reference identifying the information retrieved via the Twitter APIv2 and used to construct our variables. The information retrieved is framed in blue frames, while the variables created by using that content are indicated in red frames. The left panel shows the information extracted to create the tweet related variables; the right panel shows the information extracted to create the user-related variables. Importantly, the Twitter APIv2 is no longer available for free to researchers, although some paid solutions remain accessible.

Table C.1 and C.2 list all the variables used in our analysis. For each variable, we provide a short description, the values of its categories, scale or interval, and the data source from which it was created. More specifically, Table C.1 provides information about the outcome variables, treatment variables, and the variables capturing users' features. Some important points are noteworthy about the latter. These variables were the result of own coding, done via the OpenAI API. The input for this coding exercise is the description of the profiles provided by some of the users. It is important to mention that not all users provide a profile description. Thus, for variables like religiosity, we treat equally those who had a description but did not mention being religious and those who did not provide a description at all.

Table C.2 presents the variables capturing language metrics, valence, and emotions of text, which are used in some descriptive outputs in the paper. It also includes information on date and location. The language metrics were computed directly from the text and are used to assess differences in the features of narratives and neutral tweets. These include measures such as lexical density, complexity, and vocabulary richness. Similarly, we compute measures of valence and emotions using word lists from the NRC Dictionary Mohammad and Turney (2013). These variables allow us to examine how the emotional tone and language characteristics vary across different types of tweets.

Figure C.1: *Information Scraped via Twitter APIv2*



Notes: The figure shows two screenshots. On the left a tweet posted by the user *@SenatorReid*, on the right the same user's Twitter profile. We frame all the information retrieved via the Twitter APIv2 in blue and indicate the variable for which the information is used in red frames. In Section 3 of the paper we describe our data.

Table C.1: ***Description of Variables (Part I)***

Variable	Question/Description	Categories/Scale/Interval	Source
Outcomes			
No. of Retweets	Number of times the tweet is retweeted	$n \in [0; 29{,}526]$	Twitter APIv2
No. of Replies	Number of times the tweet is replied to	$n \in [0; 30{,}887]$	Twitter APIv2
No. of Likes	Number of times the tweet is liked	$n \in [0; 489{,}375]$	Twitter APIv2
Treatment			
Character-Role (predicted)	Detects whether the tweet contains a character-role presenting a character in a specific role	0 = not present, 1 = present	Own computation
Villain	Indicates at least one villain narrative is present in the tweet	0 = none, 1 = at least one	Own computation
Hero	Indicates at least one hero narrative is present in the tweet	0 = none, 1 = at least one	Own computation
Victim	Indicates at least one victim narrative is present in the tweet	0 = none, 1 = at least one	Own computation
Neutral	Indicates at least one narrative with neutral character representation in the tweet	0 = none, 1 = at least one	Own computation
Human	Indicates at least one human character presented in a role	0 = none, 1 = at least one	Own computation
Instrument	Indicates at least one instrument character presented in a role	0 = none, 1 = at least one	Own computation
Control Variables			
No. of Words	Number of words in the tweet (excluding mentions/hashtags)	$n \in [0; 117]$	Own computation
No. of Hashtags	Number of hashtags (#) in the tweet	$n \in [0; 26]$	Own computation
No. of Mentions	Number of mentions (@) in the tweet	$n \in [0; 51]$	Own computation
No. of Followers	Number of users following the tweet's author	$n \in [0; 133{,}245{,}480]$	Twitter APIv2
No. of Following	Number of accounts the author follows	$n \in [0; 4{,}066{,}970]$	Twitter APIv2
No. of Tweets	Total tweets produced by the author up to posting	$n \in [0; 9{,}611{,}963]$	Twitter APIv2
Author Characteristics			
Democrat	Indicates if the author's profile description identifies them as a Democrat	0 = no, 1 = yes	Own computation
Republican	Indicates if the author's profile description identifies them as a Republican	0 = no, 1 = yes	Own computation
Religious	Indicates if the author's profile description mentions a religious affiliation	0 = no, 1 = yes	Own computation
High Education	Indicates if the author's profile description mentions high educational attainment	0 = no, 1 = yes	Own computation
Children	Indicates if the author's profile description states that they have children	Count Variable for number of children	Own computation

Notes: The table contains a description of all the variables for outcomes (virality), treatment (character roles), control variables, and author characteristics. All data are sourced either from own computation or the Twitter APIv2. In Section 3 of the paper, we describe our data in more detail.

Table C.2: ***Description of Variables (Part II)***

Variable	Question/Description	Categories/Scale/Interval	Source
Quality of Text			
Lexical Density	Ratio of content words (nouns, verbs, adjectives, adverbs) to total words	$\in [0.01; 0.92]$	Own computation
Type/Token Ratio	Ratio of unique words to total words in the tweet	$\in [0; 1]$	Own computation
Reading Ease	Flesch Reading Ease measure (higher = easier to read)	$\in [1; 114.63]$	Own computation
Education needed to comprehend text	Approx. US grade level needed to understand the tweet (e.g., Flesch-Kincaid)	$\in [1.1; 54.27]$	Own computation
Emotions			
Joy	Joy is the average occurrence of words from the NRC dictionary associated with "joy" based on (Mohammad and Turney 2013).	Count variable	Own computation
Surprise	Surprise is the average occurrence of words from the NRC dictionary associated with "surprise" based on (Mohammad and Turney 2013).	Count variable	Own computation
Fear	Fear is the average occurrence of words from the NRC dictionary associated with "fear" based on (Mohammad and Turney 2013).	Count variable	Own computation
Sadness	Sadness is the average occurrence of words from the NRC dictionary associated with "sadness" based on (Mohammad and Turney 2013).	Count variable	Own computation
Anger	Anger is the average occurrence of words from the NRC dictionary associated with "Anger" based on (Mohammad and Turney 2013).	Count variable	Own computation
Other			
Date	Date of tweet creation	02.01.2010 – 25.12.2021	Twitter APIv2
Location	Highest level of precision at which the tweet could be located	{country, state (US), city}	Own computation

Notes: The table contains a description of all the variables related to the text metrics, valence, and emotions in text. In Section 3 of the paper we describe our data.

C.2 Additional Descriptive Statistics

In this section of Appendix C, we provide descriptive statistics on the observational data used to create the outputs in the paper and appendices. We begin by complementing the descriptive statistics shown in the paper with Table C.3. The tables report descriptive statistics for the set of relevant tweets, those used in the analysis. For each variable, we report the mean, median, standard deviation, minimum, and maximum value, covering the level of localization, public metrics from the user's profile, and information from the profile description.

Table C.3: ***Features of Relevant Tweets (United States, 2010-2021)***

	Mean	Median	St. Dev.	Min.	Max.
Virality					
No. of Retweets (Virality)	3.7	0	154	0	29,526
No. of Likes	19	0	1,239	0	489,375
Tweet's Characteristics					
No. of Words	29	26	13	1	64
No. of Hashtags	.47	0	1.1	0	26
No. of Mentions	1.7	1	4.2	0	51
Quality of Text					
Lexical Density	.81	.82	.081	0	1
Type/Token Ratio	.93	.94	.062	0	1
Reading Ease	56	59	22	-1148	117.67
Educa. Needed to Comprehend Text	12	11	4.8	1	54.23
Emotions					
Avg. Count of Joy Words	.048	0	.081	0	1
Avg. Count of Surprise Words	.027	0	.067	0	1
Avg. Count of Trust Words	.11	.048	.15	0	1
Avg. Count of Anger Words	.048	0	.084	0	1
Avg. Count of Disgust Words	.03	0	.07	0	1
Avg. Count of Fear Words	.15	.1	.21	0	1
Avg. Count of Sadness Words	.053	0	.093	0	1
No. of Observations	309,744				

Notes: The table displays descriptive statistics for the dataset of relevant tweets used in our analysis. We define a tweet as relevant if it features at least one character from our list. We include only character roles that appear at least 100 times, thus excluding 'US Republicans-Victim', 'Emission Pricing-Victim', 'Regulations-Victim', and 'Green Tech-Victim'. For each variable, we report the average, median, standard deviation, and minimum/maximum values. We calculate the number of words per tweet excluding hashtags and mentions. We group variables by their role in the analysis. Paper Section 3 is the reference section, where we describe and discuss the data used in the analysis.

We move on with Table C.4, showing descriptive statistics about users that posted the relevant tweets. The localization level is a dichotomous variable equal to one if the user could be located, through our geo-localization pipeline, at the state level. Roughly 93% of users could be located at least at the state level. When a user is located only at the national state, she is assigned to a 'fictitious' state called USA, also used in the stated FEs. A total of 3.4% of users' profiles were verified, at a time when verification was provided by Twitter for public figures. Users in our dataset are generally prolific: the median user has posted around 8,800 tweets (this refers to total activity since account creation, not within our dataset). In 14% of cases, users mention in their profile description that they are Democrats, and in 3.4% of cases, Republicans. Around 5% of users described themselves as religious. One quarter of users reported having higher education, and 11%

reported having children.

Table C.4: ***Characteristics of Users That Posted Relevant Tweets (United States, 2010-2021)***

	Mean	Median	St. Dev.	Min.	Max.
Localization Level					
Share State-Located	.93	1	.25	0	1
Profile's Characteristics					
Share verified	.034	0	.18	0	1
No. of Followers	8,804	434	425,031	0	133,243,353
No. of Following	1,701	668	6,381	0	841,864
No. of Tweets	25,173	8,094	58,077	1	3,671,808
Profile Description					
Share of Democrats	.14	0	.35	0	1
Share of Republicans	.034	0	.18	0	1
Share religious	.056	0	.23	0	1
Share with High Educ.	.25	0	.43	0	1
Share with Children	.11	0	.31	0	1
No. of Observations	152,560				

Notes: The table provides insights into the users' characteristics for those users that posted the relevant tweets used in our analysis. We define a tweet as relevant if it features at least one character from our list. We include only character roles that appear at least 100 times, thus excluding 'US Republicans-Victim', 'Emission Pricing-Victim', 'Regulations-Victim', and 'Green Tech-Victim'. For each variable, we report the average, median, standard deviation, and minimum/maximum values. We calculate the number of words per tweet excluding hashtags and mentions. We group variables by their role in the analysis. Paper Section 3 is the reference section, where we describe and discuss the data used in the analysis.

Table C.4 provides insights into the characteristics of relevant tweets, defined as those containing at least one of the characters of interest in this analysis. For comparison, Table C.5 reports descriptive statistics for all tweets classified through our GPT pipeline. This broader dataset includes the relevant tweets used in our analysis, tweets classified as addressing climate change policy without mentioning any of our characters, tweets discussing the existence of man-made climate change more generally, and tweets not related to climate change at all.

Comparing the paper Table C.3 and Table C.5 some noteworthy points emerge. The subset of relevant tweets, compared to the totality of tweets, is generally more viral with retweets and like being almost twice as high. While relevant tweets are generally longer, the amount of hashtags and mentions used is comparable. Virtually all measures of text quality, emotions, and valence in text are comparable between the two datasets. Overall, the relevant tweets tend to be longer and more viral.

Table C.5: ***Features of All Tweets (United States, 2010-2021)***

	Mean	Median	St. Dev.	Min.	Max.
Virality					
No. of Retweets (Virality)	2.3	0	269	0	194,217
No. of Likes	11	0	1,229	0	896,759
Tweet's Characteristics					
No. of Words	23	20	13	0	65
No. of Hashtags	.43	0	1.1	0	28
No. of Mentions	1.6	1	4.9	0	51
Quality of Text					
Lexical Density	.79	.81	.098	0	1
Type/Token Ratio	.94	.95	.062	0	1
Reading Ease	60	64	24	-2840	120.21
Educa. Needed to Comprehend Text	10	9.6	5.1	0	280.4
Emotions					
Avg. Count of Joy Words	.04	0	.082	0	1
Avg. Count of Surprise Words	.026	0	.077	0	1
Avg. Count of Trust Words	.096	0	.16	0	1
Avg. Count of Anger Words	.048	0	.1	0	1
Avg. Count of Disgust Words	.031	0	.076	0	1
Avg. Count of Fear Words	.17	.067	.25	0	1
Avg. Count of Sadness Words	.045	0	.09	0	1
No. of Observations	1,151,671				

Notes: The table reports descriptive statistics for the dataset of all tweets used in the GPT classification. For each variable, we report the average, median, standard deviation, and minimum/maximum values. We include only character roles that appear at least 100 times, thus excluding 'US Republicans-Victim', 'Emission Pricing-Victim', 'Regulations-Victim', and 'Green Tech-Victim'. We calculate the number of words per tweet excluding hashtags and mentions. We group variables by their role in the analysis. Paper Section 3 is the reference section, where we describe and discuss the data used in the analysis.

C.3 Distribution of Narratives and Virality Outcomes

In this subsection of Appendix C, we provide additional output describing the distribution of political narratives and the virality outcomes used in our analysis. As a first step, we show how tweets were classified by GPT into the different categories defined in the two steps of our pipeline, in Figure C.2. As explained in Appendix A, in the first part of the pipeline GPT classifies tweets into four categories: *Irrelevant* (not about climate change), *Assert* (stating that man-made climate change exists), *Deny* (denying man-made climate change exists), and *Policy* (tweets about climate change policy).

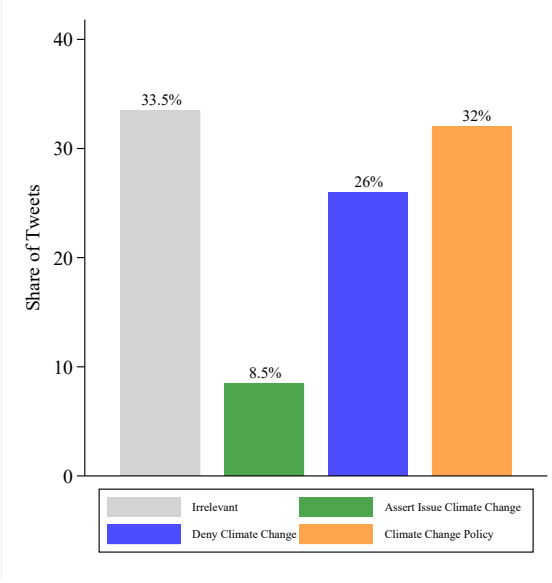
Figure C.2a shows the distribution of tweets classified in the first step of our pipeline. Roughly 33% of tweets are irrelevant, meaning they are either too unclear to classify, too short, or unrelated

to climate change despite mentioning keywords from our list. Tweets classified as discussing climate change policy make up about 32% of all tweets. A large share of the conversation on Twitter/X revolves around simply asserting or denying the existence of man-made climate change. These tweets do not contribute to the policy debate but instead reflect fixed and polarized positions on either side of the broader discussion.

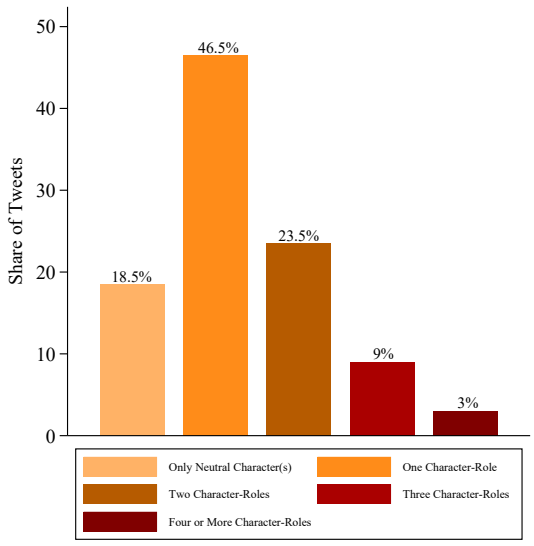
Only for tweets in the *Climate Change Policy* category do we prompt GPT to identify characters of interest. Those tweets that feature at least one character from our list, form what we call the set of `relevant` tweets: a total of 309,744 tweets. Figure C.2b provides a breakdown of the set of `relevant` tweets, where a clear picture emerges. In 18.5% of tweets, characters are present but only in neutral form, meaning they are not assigned one of the three roles from the drama triangle. This subset serves as the comparison group in most of our analysis and forms the basis for estimating the effect of *Political Narratives*. About 46% of tweets contain at least one character-role, framing a character as a hero, villain, or victim. Two character-roles appear together in 23% of tweets, three in 9%, and four or more in 3%.

Figure C.2: ***Total Share of Tweets in GPT's Classification***

(a) Distribution of Policy Related and Other Tweets



(b) Total Share of Narratives by Presence of Characters



Notes: The figures provide insights into the classification obtained through our pipeline. On the left, Figure C.2a reports results from the first step of the classification, showing the share of tweets originating from the US that were labeled as irrelevant to the climate change discussion (gray), simply asserting the importance of the matter (green), denying the matter (blue), or focusing on climate change narratives about policies and solutions. On the right, Figure C.2b focuses only on the latter group of policy-related tweets. For these tweets, we show the distribution by character framing: the share featuring characters only in a neutral way, the share featuring one character-role, two character-roles, three, and four or more.

In the second part of this section, we provide additional details into the distribution of our character-roles. In other words, we dive into the classification of characters, divided into roles and

neutral framing. We aim to complement the paper Table 2, which displays the distribution in shares, excluding those character-roles that did not reach at least 100 occurrences in the full set of classified tweets. We report the distributions in Table C.6.

Analyzing Table C.6, several key points stand out. As mentioned in the paper and above, some character-roles did not reach 100 occurrences: US Republicans–Victim (95), Emission Pricing–Victim (11), Regulations–Victim (29), and Green Tech–Victim (88). Although the 100-occurrence threshold is discretionary, we consider it a reasonable cutoff. We exclude character-roles below this threshold from the analysis by dropping their columns from the dataset.

Corporations and US People are the most common characters, each appearing in roughly 100,000 instances. However, many of these appear in neutral framing. The two most frequent character-roles are Fossil Industry–Villain and Green Tech–Hero, which dominate much of the public discourse. Overall, victim narratives are the least common. Only Developing Economies and US People are often framed as victims, with 5,308 and 18,113 occurrences, respectively.

Table C.6: ***Frequency of Character-Roles in Relevant Tweets (United States, 2010-2021)***

Panel A: Human Characters

	Hero	Villain	Victim	Neutral	Total
Developing Economies	777	7,025	5,308	1,166	14,276
US Democrats	33,450	10,648	333	7,924	52,355
US Republicans	683	55,568	95	9,261	65,607
Corporations	5,767	47,770	375	46,180	100,092
US People	23,450	3,221	18,113	56,833	101,617

Panel B: Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emission Pricing	11,955	7,337	11	6,985	26,288
Regulations	20,420	7,968	29	18,656	47,073
Fossil Industry	366	69,009	238	9,423	79,036
Green Tech	67,621	4,911	88	18,737	91,357
Nuclear Tech	5,077	1,879	107	2,577	9,640

Notes: The table displays the absolute frequencies of character-roles in the classified data. Sums are computed using the dataset of relevant tweets, used in our analysis. We define a tweet as relevant if it features at least one character from our list. Generally, in the analysis we perform in the paper, we exclude the character-roles that do not appear at least 100 times, thus excluding 'US Republicans-Victim', 'Emission Pricing-Victim', 'Regulations-Victim', and 'Green Tech-Victim', nevertheless we report their frequency in the table. Panel (a) displays the sums for characters of the human type, while Panel (b) displays the same for characters of the instrument type. The column Neutral in both panels reports cases where the character is present in the tweet but is not depicted in one of the three specific roles. The occurrence of character-roles is not mutually exclusive, meaning multiple roles may appear in the same tweet. The main paper Table 2 displays similar information provide insights into shares, computed excluding the categories that do not reach 100 instances.

We provide detailed info about the evolution of the character roles over time in Table C.7. These

statistics complement the information in Figure 3 by displaying the average percentage share for each character-role over time, along with the standard deviation, and the minimum and maximum values.

Table C.7: ***Share of Character-role Combinations over Time***

	Mean	St. Dev.	Min.	Max.
Green Tech-Hero	23.05	9.82	10.42	42.15
Fossil Industry-Villain	16.93	2.73	10.97	20.64
U.S. Republicans-Villain	11.80	5.83	2.79	23.63
Corporations-Villain	10.65	2.12	6.84	14.96
U.S. Democrats-Hero	6.69	2.80	3.20	12.83
U.S. People-Hero	5.19	1.18	3.54	7.99
Regulations-Hero	4.56	1.10	2.82	6.24
U.S. People-Victim	3.41	1.57	1.27	6.08
Emiss. Pricing-Hero	3.24	0.79	2.09	4.63
Emiss. Pricing-Villain	2.16	0.74	1.20	3.72
Corporations-Hero	2.06	1.88	0.90	7.82
U.S. Democrats-Villain	1.93	0.97	0.79	3.97
Regulations-Villain	1.75	0.36	1.28	2.38
Developing Economies-Villain	1.43	0.46	0.75	2.27
Developing Economies-Victim	1.21	0.28	0.83	1.77
Green Tech-Villain	1.19	0.38	0.66	1.85
Nuclear Tech-Hero	0.93	0.42	0.29	1.66
U.S. People-Villain	0.61	0.26	0.23	1.05
Nuclear Tech-Villain	0.56	0.37	0.34	1.70
Developing Economies-Hero	0.22	0.08	0.09	0.39
U.S. Republicans-Hero	0.13	0.08	0.01	0.32
Corporations-Victim	0.10	0.07	0.00	0.27
Fossil Industry-Hero	0.07	0.04	0.01	0.13
U.S. Democrats-Victim	0.06	0.04	0.01	0.14
Fossil Industry-Victim	0.05	0.02	0.02	0.09
Nuclear Tech-Victim	0.02	0.01	0.00	0.05

Notes: The table reports descriptive statistics for the share of each character-role over time. For each character-role, we report the average, standard deviation, and minimum/maximum values. We include only character roles that appear at least 100 times, thus excluding 'US Republicans-Victim', 'Emission Pricing-Victim', 'Regulations-Victim', and 'Green Tech-Victim'. Statistics must be interpreted as percentages. Paper Subsection 4.3 is the reference section, where we describe and discuss the evolution of character-role combinations over time.

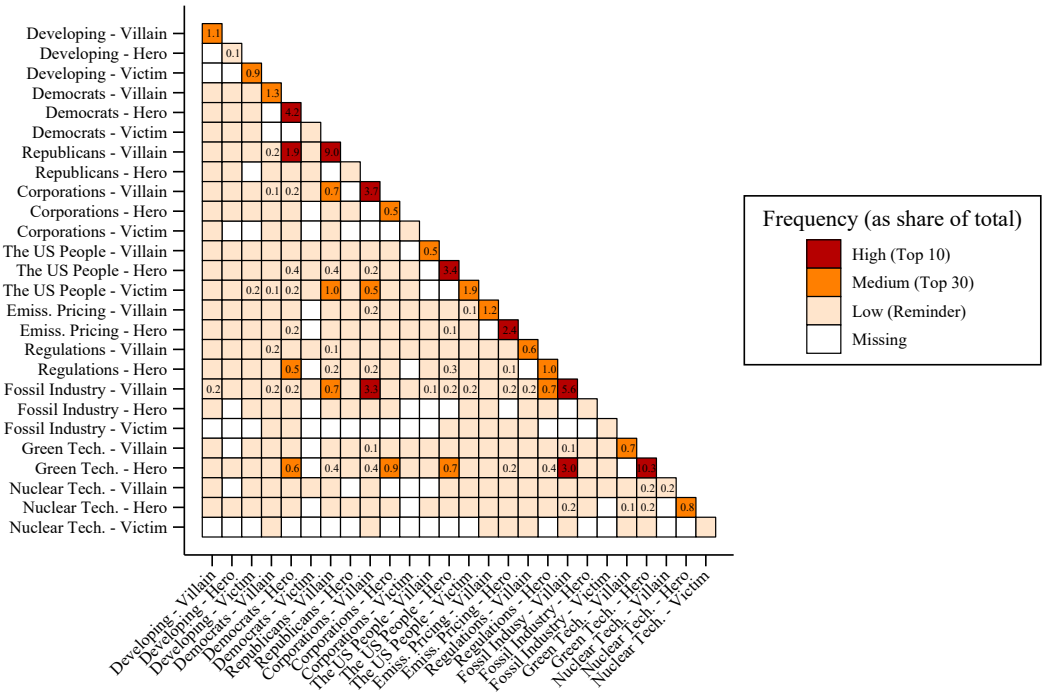
Below, we complement the information reported in the paper Figure 4 and Figure 5. Figure C.3 and Figure C.4 show the same plots, but complemented by annotating the most frequent single character-roles – in the diagonal – or configurations of two character-roles – off the diagonal. Besides the additional information on the graphs, we want to provide more detail on the computation of these figures.

For what concerns Figure C.3, we restrict to the subset of relevant tweets containing one or two character–roles, denoted $\mathcal{D}^{\text{rel}}_{[1,2]} = \{\, i \in \mathcal{D}^{\text{rel}} : 1 \leq N_i \leq 2 \,\}$, where the number of (non-neutral)

character–roles in tweet i is defined as $N_i = \sum_{k \in \mathcal{K}} \sum_{r \in \mathcal{R}} r_{ikr}$. We then compute a symmetric matrix m_{ij}, where each index i and j corresponds to a character–role combination. The diagonal entries m_{ii} represent the share of tweets in $\mathcal{D}^{\text{rel}}_{[1,2]}$ that contain only character–role i, while the off-diagonal entries m_{ij} for $i \neq j$ capture the share of tweets that contain exactly the pair (i, j). Each matrix entry is computed as $m_{ij} = \frac{1}{|\mathcal{D}^{\text{rel}}_{[1,2]}|} \sum_{t \in \mathcal{D}^{\text{rel}}_{[1,2]}} \mathbb{1}(i \in t \wedge j \in t \wedge N_t = |\{i, j\}|)$, where $\mathbb{1}(\cdot)$ is the indicator function and $\{i, j\}$ is the set of character–roles considered. To aid interpretation, we highlight the Top 10 and Top 30 most frequent configurations in different colors.

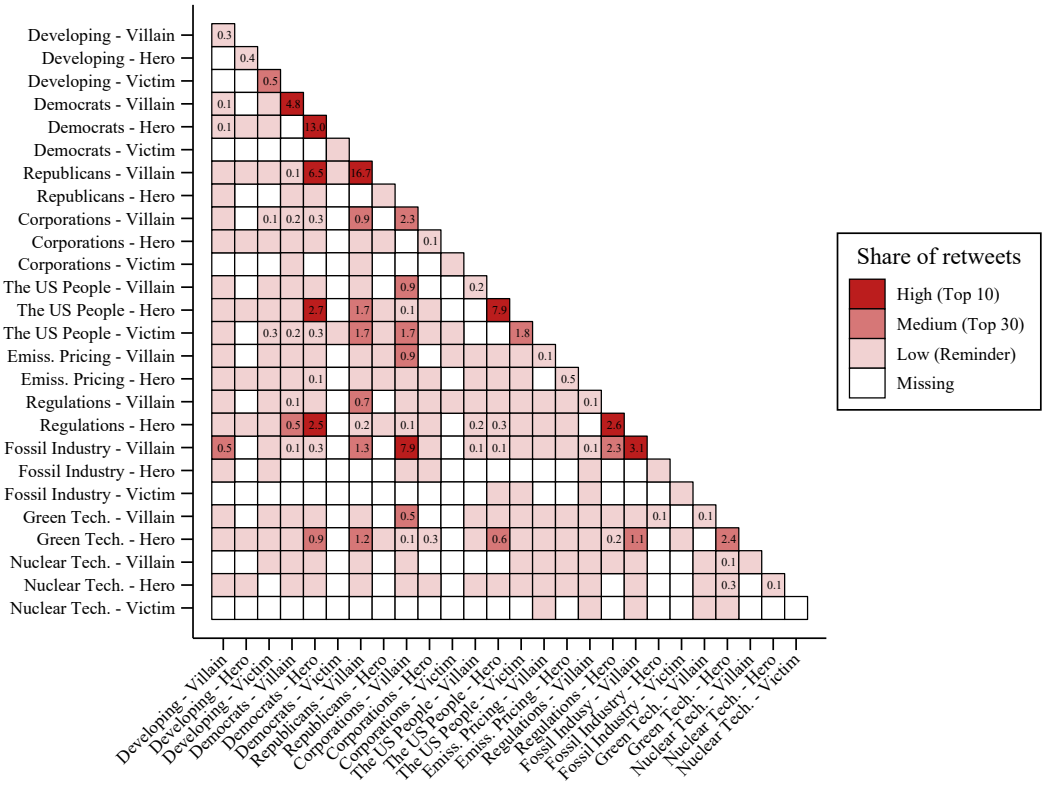
For what concerns Figure C.4, we follow the same logic but replace tweet counts with retweet counts. That is, we compute a weighted matrix $m^{(w)}_{ij}$ where each entry reflects the share of total retweets received by tweets in $\mathcal{D}^{\text{rel}}_{[1,2]}$ that contain a given character–role configuration. Let $w_t \geq 0$ denote the number of retweets received by tweet t. Then, for any pair of character–roles i and j, the weighted entry is defined as $m^{(w)}_{ij} = \frac{1}{\sum_{t \in \mathcal{D}^{\text{rel}}_{[1,2]}} w_t} \sum_{t \in \mathcal{D}^{\text{rel}}_{[1,2]}} w_t \cdot \mathbb{1}(i \in t \wedge j \in t \wedge N_t = |\{i, j\}|)$. As before, diagonal entries $m^{(w)}_{ii}$ correspond to tweets featuring only character–role i, while off-diagonal entries $m^{(w)}_{ij}$ for $i \neq j$ capture tweets featuring exactly the pair (i, j). The matrix thus describes how the total number of retweets is distributed across different narrative structures, without adjusting for how frequently they occur. To aid interpretation, we again highlight the Top 10 and Top 30 configurations by total retweet share using different colors.

Figure C.3: ***Absolute Frequency of Character-Roles Combinations - Tweets with One or Two Character-Roles***



Notes: The figure shows the frequency of each character–role appearing alone or in combination with another character–role divided by all political narrative tweets with one or two character–roles. The diagonal of the matrix shows how often each character-role appears alone in a tweet. Tweets with three or more character-roles are excluded. We report character-roles that appear at least 100 times over 2010-2021; excluded characters are US Republicans-Victim, Emission Pricing-Victim, Regulations-Victim, and Green Tech-Victim. We avoid clutter, we use a color scheme to highlight the top 10 most frequent character-role combinations, the rest of the top 30, and the remaining pairs. White indicates a pair that never appears together. The top 10 are in order: Green Tech-Hero (10.27%), US Republicans-Villain (8.95%), Fossil Industry-Villain (5.56%), US Democrats-Hero (4.21%), Corporations-Villain (3.66%), US People-Hero (3.40%), Fossil Industry-Villain + Corporations-Villain (3.27%), Green Tech-Hero + Fossil Industry-Villain (3.02%), Emission Pricing-Hero (2.40%), US Republicans-Villain + US Democrats-Hero (1.92%).

Figure C.4: ***Virality of Character-Roles - Tweets with One or Two Character-Roles***



Notes: The figure shows the retweet share of each character–role appearing either alone or in combination with another role, among relevant tweets containing one or two roles. Retweet rates are computed as the share of total retweets received by a given role (or pair) relative to all retweets of tweets with one or two roles. The diagonal of the matrix shows the retweet rate when each character-role appears alone. Tweets with three or more character-roles are excluded. We report character-roles that appear at least 100 times over 2010-2021; excluded characters are US Republicans-Victim, Emission Pricing-Victim, Regulations-Victim, and Green Tech-Victim. To avoid visual overload, we do not display exact rates. Instead, we use a color scheme to highlight the top 10 most frequently retweeted character-role combinations, the top 30 (which includes the top 10), and the remaining pairs. White indicates a pair that never appears together. The top 10 in order is: US Republicans-Villain (16.72%), US Democrats-Hero (12.98%), US People-Hero (7.93%), Fossil Industry-Villain + Corporations-Villain (7.89%), US Republicans-Villain + US Democrats-Hero (6.51%), US Democrats-Villain (4.80%), Fossil Industry-Villain (3.08%), US People-Hero + US Democrats-Hero (2.71%), Regulations-Hero (2.58%), Regulations-Hero + US Democrats-Hero (2.52%).

D Robustness Checks: Observational Data

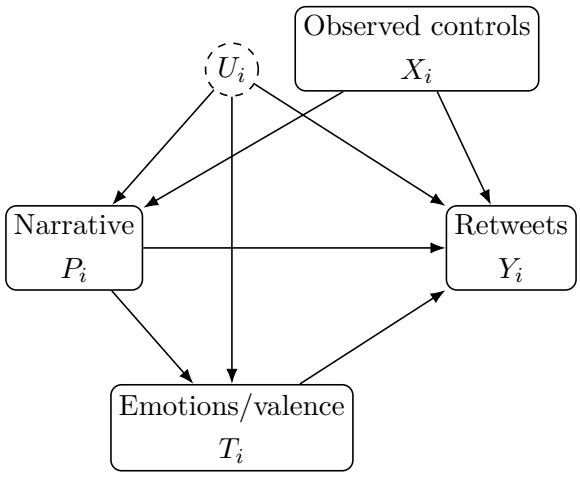
D.1 DAG Depiction of Language Metrics and Emotions/Valence Controls

DAG interpretation. Consider the DAG with nodes P_i (narrative), T_i (emotions/valence), Y_i (retweets), X_i (observed author/tweet controls), and U_i (unobserved shocks, e.g., community-level salience, off-platform events). The causal structure is:

$$P_i \to T_i \to Y_i, \qquad P_i \to Y_i, \qquad X_i \to P_i,\ X_i \to Y_i, \qquad U_i \to P_i,\ U_i \to Y_i,\ U_i \to T_i.$$

Backdoor paths from P_i to Y_i run through (X_i, U_i). We block X_i paths by conditioning on X_i and fixed effects; U_i is unobserved and remains a residual concern. Crucially, T_i is *not* a confounder but rather a *mediator*. *Baseline:* not conditioning on T_i leaves the causal paths $P_i \to Y_i$ and $P_i \to T_i \to Y_i$ intact (total association). *Conditioning on T_i:* (i) blocks the mediated path $P_i \to T_i \to Y_i$; (ii) can introduce *collider bias* if $U_i \to T_i$ and $P_i \to T_i$, because T_i becomes a collider on $P_i \to T_i \leftarrow U_i \to Y_i$; conditioning on T_i then opens the $P_i \leftrightsquigarrow U_i \to Y_i$ path. Hence, the "mechanism-adjusted" specification is informative only under strong conditions (no unobserved U_i that jointly affects T_i and Y_i) and should not be interpreted as the full direct effect.

Figure D.1: ***External mediator and controls.***



Notes: T_i is drawn below the main path ($P_i \to Y_i$) to emphasize its mediating role; X_i is drawn externally on the upper right, feeding into both P_i and Y_i; U_i is unobserved and influences all three.

D.2 Heterogeneity by Number of Followers

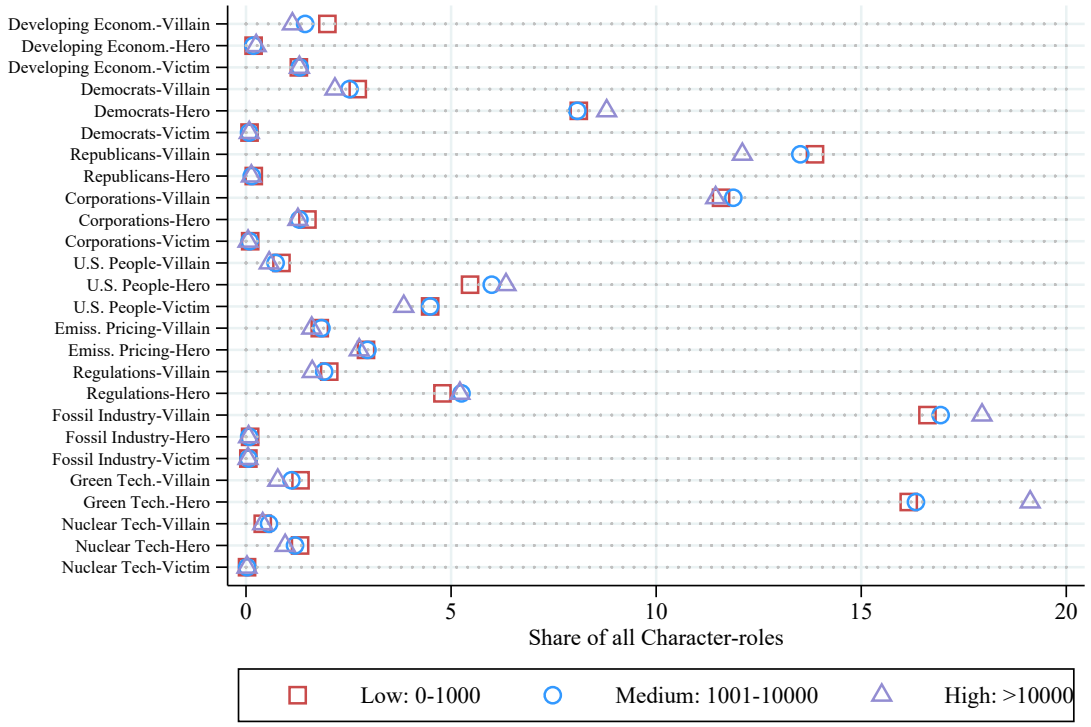
In this section of Appendix D, we tackle the question: How does the users' reach within the platform – expressed as the number of followers – impact the kind of narratives the users spread and the virality of these narratives?

We measure the reach of a profile by the number of its followers. We divide users into three groups: low reach (0 to 1,000 followers), medium reach (1,001 to 10,000 followers), and high reach

(more than 10,000 followers). As a first step, we ask whether the types of narratives differ across users with different reach levels. This check ensures that narrative content is comparable across reach categories. Figure D.2 plots the share of each character-role relative to all character-roles within each reach group. Squares represent low-reach profiles, circles medium-reach, and triangles high-reach. We plot the shares for all character-roles that appear at least 100 times in the full dataset.

Analyzing Figure D.2, it appears clear that there are virtually no differences in narrative content across profile categories, with only a few exceptions. Low and medium reach profiles show extremely similar patterns. High-reach profiles differ in four cases: they post a slightly higher share of US Democrats–Hero narratives, a considerably lower share of US Republicans–Villain narratives, a higher share of Fossil Industry–Villain narratives, and a considerably higher share of Green Tech–Hero narratives. The latter is also the most widespread narrative used by high-reach users. Overall, differences across categories are rare, except that high-reach users appear slightly less politicized and more concerned with energy sources. A question remains: Does this lack of difference also translate into a similar impact of featuring narratives on the virality of their tweets?

Figure D.2: ***Share of Character-Roles by Reach of the Profile***

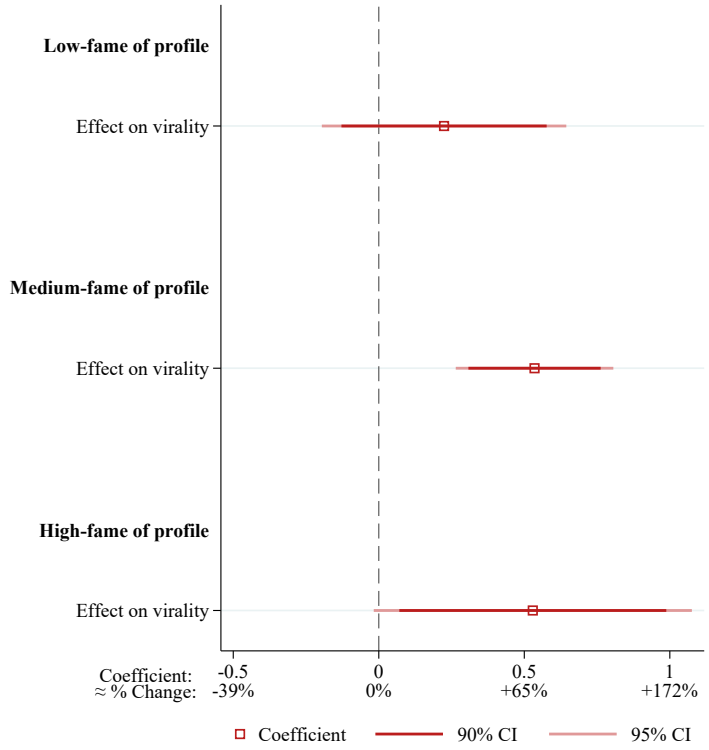


Notes: The figure shows the share of character-roles featured in relevant tweets by profile reach. Squares represent profiles with fewer than 1,000 followers, circles represent those with 1,001–10,000 followers, and triangles represent profiles with more than 10,000 followers. Shares are computed by dividing the number of times each character-role appears in tweets from a given profile category by the total number of character-roles used within that category.

After showing that narrative content is similar across profiles with different reach levels, we turn to the next question: Does this similarity lead to a similar impact of narratives on the virality of tweets across these groups? In other words, building on the main empirical results of the paper – that show narratives increase the virality of tweets compared to neutral framing of the same characters – we now ask whether this effect holds within different types of profiles. Figure D.3 provides insights on this question. The coefficients plot shows the impact of containing a narrative on virality, the count of retweets, compared to featuring characters in neutral framing. The results are based on a slightly modified version of the main specification used in the paper. We use a Poisson Pseudo-Maximum Likelihood regression model, including character fixed effects, and hour, week, and year-state fixed effects. We do not include author characteristics to avoid capturing variation that defines the reach categories themselves.

This exercise provides some interesting results. Despite larger confidence intervals for the high-reach group – due to a smaller number of users – medium- and high-reach profiles show similar patterns. In both groups, narratives increase virality in line with the main results of the paper. In contrast, tweets from low reach profiles show no measurable impact of narratives on virality. This finding lends itself to several interpretations. The most likely is that at low levels of reach, the content posted by these users rarely goes viral, regardless of whether it contains a narrative or presents neutral framing. In other words, at low follower counts, factors other than content – such as limited reach or engagement – may prevent virality, making the narrative effect negligible.

Figure D.3: ***Regression Results - Impact of Political Narratives on Virality by Reach of Profile***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regression models testing the effect of featuring at least one character-role vs. featuring characters only in a neutral role on virality, measured as the count of retweets. We explore the effects in three sub-populations: top panel, for profiles with fewer than 1,000 followers, middle panel for users those with 1,001–10,000 followers, and bottom panel for profiles with more than 10,000 followers. The x-axis reports coefficient estimates, 90% confidence intervals (dark red), and 95% confidence intervals (light red). We label also the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. All regressions control for author characteristics (verified status, number of followers/followings, total tweets created, party affiliation, religiosity, higher education, and parenthood) and include character fixed effects. We also include hour, week-of-year, and year–state fixed effects. Standard errors are clustered at the week level, covering the full time frame.

A note of caution is necessary for this analysis. The information on follower count reflects the number of followers at the time of data extraction, which means it is an ex-post measure relative to when the tweets were posted. It is therefore possible that the reach of some profiles results from the use of narratives, rather than causing it. While this is the best available approach given the data, the results should be interpreted carefully.

D.3 Impact of Major Twitter Algorithm Changes

In the paper, we show that featuring characters in political narratives generates a virality premium compared to presenting the same characters in a neutral way. This effect remains strong and consistent even after introducing strict controls for time, user profiles, and fixed effects. Since we study a social media platform, another crucial mechanism shaping virality is the algorithm that

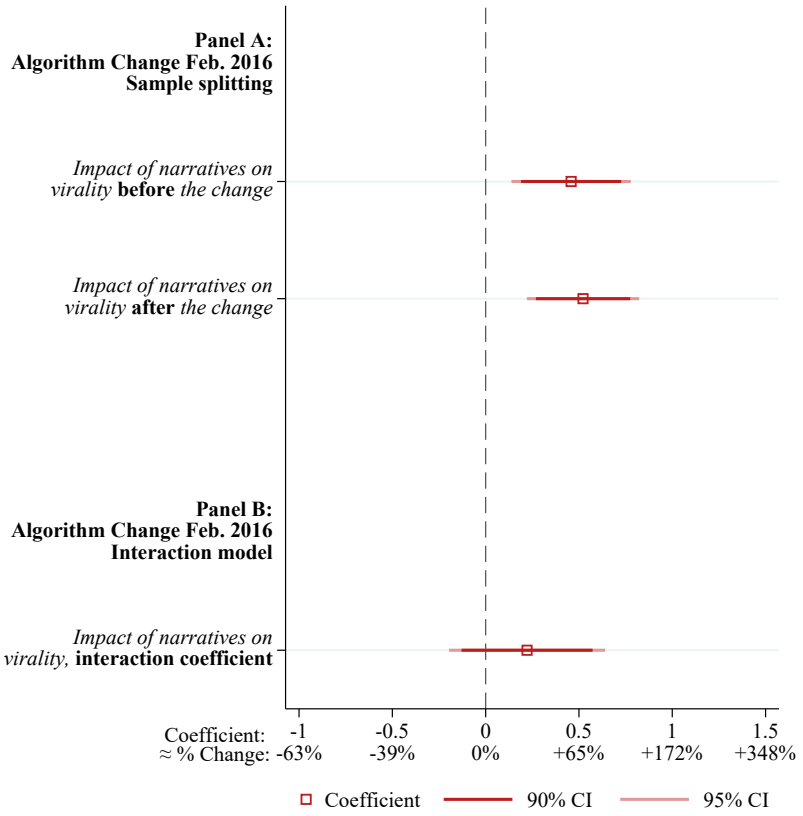
structures content exposure by curating each user's timeline. In this section, we investigate the role of Twitter's algorithm, exploiting a major change that occurred during our study period.

A social media platform's timeline algorithm can be understood as a set of rules and mechanisms to interpret signals that determines what content a user sees when using the platform. The term timeline reflects the original mechanism used to regulate this flow in most of social media platforms: posts were ranked chronologically, with users shown the most recent content from the accounts they followed. Over time, however, this purely time-based ranking was phased out and replaced with what is commonly called the algorithmic timeline. The idea is straightforward: an AI-powered algorithm curates each user's feed, tailoring it to individual tastes and past behavior, and prioritizing the content the system predicts the user is most likely to engage with.

Twitter adopted this shift in February 2016, when it introduced its own timeline algorithm. As BuzzFeed reported at the time, the algorithm was designed as "a way for Twitter to elevate popular content, and could solve some of Twitter's signal-to-noise problems. [...] The timeline will reorder tweets based on what Twitter's algorithm thinks people most want to see, a departure from the current feed's reverse chronological order" (BuzzFeed). In many ways, this marked a before-and-after moment for the platform, fundamentally reshaping how information and content were consumed. Some described it as "arguably the most fundamental change it has ever made: a major tweak to the timeline" (WIRED). Others were more skeptical. As VICE put it, "2016 was the year of politicians telling us what we should believe, but it was also the year of machines telling us what we should want" (VICE).

In the context of our study, the transition from a chronological to an algorithmic timeline could be a decisive factor shaping how narratives spread online. We do not have a clear prior on the potential effects of this change. Algorithms are designed to maximize engagement by showing users content they are most likely to interact with, based on past behavior and users' characteristics. This mechanism could amplify the virality of narratives if their emotional and dramatic structure makes them especially engaging. At the same time, algorithmic curation may work in the opposite direction: by reducing the disproportionate visibility of a few highly prolific accounts, it could dilute the dominance of narrative-heavy users and favor a more balanced and maybe neutral flow of information. Finally, it is possible that the algorithm change had little effect on our outcome of interest. If narratives are intrinsically more viral than neutral content, their relative advantage might persist regardless of how the platform ranks posts. Since the internal workings of the algorithm are not clear to us, the net effect is ultimately an empirical question, which we explore in Figure D.4.

Figure D.4: ***Impact of the Main Algorithm Change in Twitter History***



Notes: The figure shows coefficients from Poisson Pseudo-Maximum Likelihood regressions testing the effect of featuring at least one character-role, compared to featuring only neutral characters, on virality (measured as the number of retweets). The analysis focuses on Twitter's major algorithmic change in February 2016, when the platform moved from a chronological timeline to an algorithm-based feed, where an algorithm ranked and prioritized content for each user. The two panels correspond to different specifications: the top panel shows the impact of *Political Narratives* on virality before (top coefficient) and after the algorithm change (second coefficient), and the bottom panel reports the interaction effect between narratives and a post-change indicator variable. The x-axis reports coefficient estimates and the corresponding approximate percentage change, computed as $e^{\beta} - 1$ and rounded to the nearest unit. All regressions control for author characteristics (verified status, followers, followings, total tweets created, party affiliation, religiosity, higher education, parenthood status) and include character, hour, week-of-year, and year–state fixed effects. Standard errors are clustered at the week level.

In Figure D.4, we examine the impact of Twitter's algorithmic change in two ways. Panel A compares the relationship between narratives and virality before and after the change using our standard specification: a Poisson Pseudo-Maximum Likelihood regression with user controls, character fixed effects, and a full set of time fixed effects. Panel B shows results from a specification that includes an interaction term between the narrative indicator and a post-change dummy (equal to 1 after the algorithm switch and 0 before). While we are aware that interaction terms in Poisson models require cautious interpretation, our goal here is simply to test whether the interaction effect is significantly different from zero.

The figure provides a clear message. Splitting the sample shows consistent evidence of the

impact of narratives on virality: both before and after the algorithm change, the effect is positive, statistically significant, and comparable in size to the estimates over the full time period presented in the main paper. The interaction term, while positive, is not statistically significant. Taken together, the results suggest that the algorithm change had little effect on the relationship between narratives and virality. Although this exercise has clear limitations and should be interpreted with caution, it is nonetheless striking that the narrative premium remains virtually unchanged, pointing to a dynamic that may go beyond the algorithmic structure of the platform.

D.4 Output Excluding Potential Bots

In this section of Appendix D, we provide an additional robustness check on the main results of the paper about the virality of political narratives. In particular, we address the important issue of bot activity on the social media platform Twitter/X. In this context, a bot can be defined as an account operated by an algorithm, programmed to automate actions such as generating content, liking, retweeting, and commenting on other users' posts. Due to the automated nature of their behavior, bots are capable of interacting with a large number of users and can, at times, achieve considerable visibility. Given this potential influence, we test whether our results on the determinants of virality are affected by the presence and activity of bots.

Identifying bots on social media is challenging, as there is no universally accepted definition or detection method. Tools such as Botometer offer sophisticated ways to classify accounts as either bots or humans. However, these tools require a substantial number of tweets per user to be effective, which is a limitation in our case due to data constraints. Furthermore, the Twitter/X API is no longer accessible, preventing us from employing such tools. We therefore adopt a simpler and more practical approach tailored to our dataset.

We implement two complementary strategies to detect potential bots based on both tweet content and user characteristics. Specifically, we define tweets as originating from potential bots if they meet one or both of the following two conditions:

1. **Repeated identical text**: Within our dataset, we observe instances where identical tweets appear multiple times. These are exact text duplicates, yet each is assigned a distinct tweet ID by the Twitter/X API, confirming that they are separate posts. We classify a tweet as bot-generated if its text appears at least five times, whether posted by the same author or by different authors. We interpret such repetitions as attempts to amplify specific narratives or content.

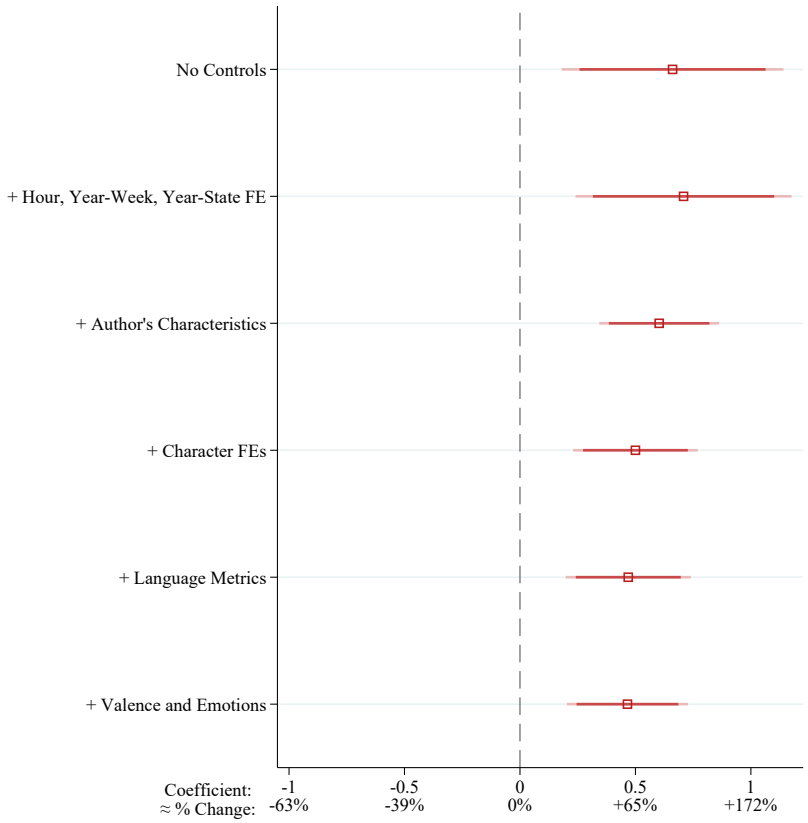
2. **Authors' activity patterns**: Following Chu et al. (2012), we use the *reputation* metric, calculated as the number of followers divided by the sum of followers and followees of a user. Bots typically have low reputation, as they tend to follow many accounts indiscriminately. Additionally, we incorporate insights from Tabassum et al. (2023), who highlight the unusually high activity levels of bots. Specifically, bots tend to produce an exceptionally large number of tweets. Combining these two indicators, we flag tweets as bot-generated if their authors

fall in the bottom 25% of the reputation distribution and simultaneously in the top 25% of the distribution of total tweets produced. The threshold of 25% is purely discretionary, but we argue it is a fairly conservative choice.

There is little overlap between the two definitions, hopefully indicating that we capture different kinds of bots successfully. Among our relevant tweets, a total of 16,700 tweets were identified through definition 1., a total of 5,748 through definition 2., while only 115 overlap. We exclude these tweets when reproducing Figure 7, which presents the main results of the paper. Figure D.5 plots the coefficients from the same models used in the paper, but excludes tweets potentially posted by bots. The models apply increasingly restrictive specifications, moving from the baseline in the top panel to the most controlled model in the bottom panel. The results remain virtually unchanged, suggesting that bots do not play a central role in shaping the discussion on climate change policy, at least within the scope of our dataset.

In conclusion, this section provides evidence that bots have limited influence on the conversation about climate change policy. However, these results should be interpreted with caution and should not be taken to mean that bots are not important for shaping social media discussions more broadly. First, in recent years, significant improvements in bot programming may have enhanced their performance in ways not captured during our study period. Second, while climate change policy does not appear to be heavily affected, other topics – such as war, abortion, or general politics – may be much more vulnerable to bot activity. Finally, as noted above, there is no exact method for identifying bots, and the approach we use may have limitations in accurately detecting automated accounts.

Figure D.5: ***Regression Results - Impact of Political Narratives on Virality, Excluding Potential Bots***

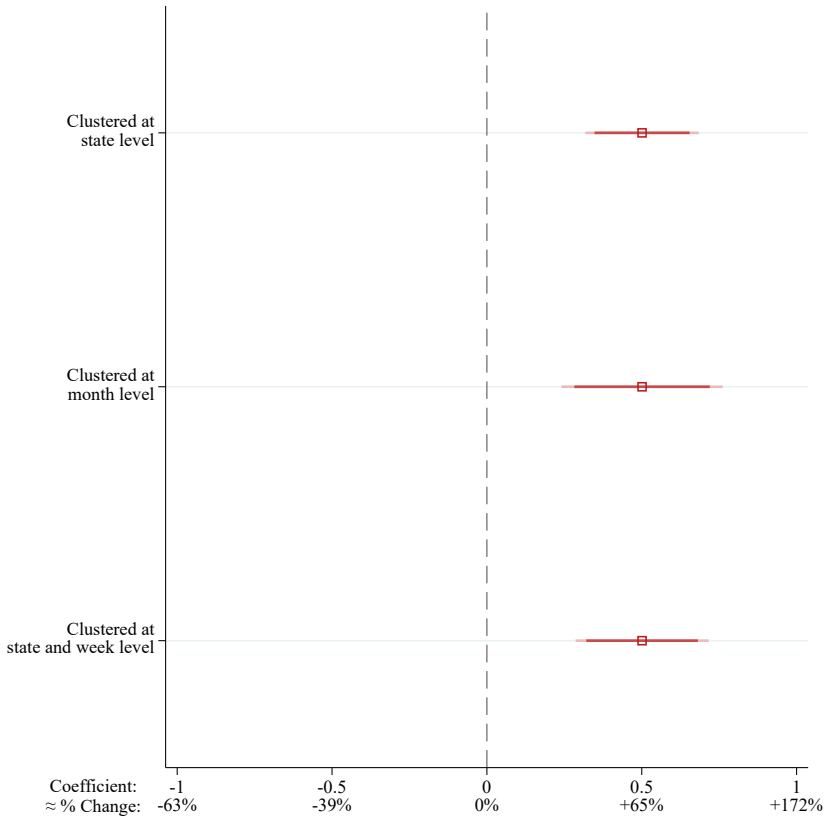


Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regression models testing the effect of featuring at least one character-role vs. featuring characters only in a neutral role on virality, measured as the count of retweets. For the analysis we exclude tweets potentially produced by bots, where bots are defined as explained in Appendix Subsection D.4. The x-axis reports coefficient estimates along with the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. Panels display results from increasingly restrictive models. The first model includes only the indicator variable for containing a character-role and clusters standard errors at the week level. The second model adds hour, week of the year, and year-state fixed effects. The third model controls for author characteristics: verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status. The fourth model adds character fixed effects. The fifth and sixth models include, respectively, language metrics and valence/emotions.

D.5 Alternative Standard Errors Clustering

In this section, we reproduce the results of Figure 7 clustering standard errors at different levels. In Figure D.6, we take our main specification model and use different clusters of the standard errors. As a reminder, our main specification includes hours of the day, week of the year times year, and year times state FEs. It also includes character fixed effects and authors' characteristics. In the paper, the model clusters standard errors at the week level. In the figure, we cluster at the state level, the monthly level, and the state + week level. The results of the analysis are robust to different clustering specifications, as can be seen in the three coefficients in the graph. As it is clear, the results are robust to different clustering.

Figure D.6: ***Regression Results - Impact of Political Narratives on Virality - Alternative SE Clustering***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regression models testing the effect of featuring at least one character-role vs. featuring characters only in a neutral role on virality, measured as the count of retweets. The x-axis reports coefficient estimates along with the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. Panels display results from the main specification model which includes hour, week of the year, year-state fixed effects, author characteristics, and character fixed effects. For each model we cluster standard errors at a different level. The first model has state level clustering. The second model clusters standard errors at the monthly level. The third model clusters standard errors at the state + week level. Figure 7 in the paper is the reference plot.

D.6 Alternative Outcomes and Estimation Methods

In this section, we want to dive deeper into the choice of model specifications used in our analysis. We explain the motivations and reasoning behind our decisions, while also considering potential alternative approaches. Our modeling choices are driven by the particular nature of the data and the research questions we address. In particular, the outcome variables of our analysis – retweets in the paper, likes and replies in the appendix – are skewed count variables: many observations take low values (including zeros), while a few take extremely high values. Such distributions are common in social media engagement and in other domains shaped by self-reinforcing processes. They pose unique challenges and require careful consideration when selecting an appropriate modeling strategy. In similar cases, researchers often apply a log transformation to ease interpretation and reduce skewness before estimating models with Ordinary Least Squares (OLS). This would be a

sensible solution if the dependent variables were strictly positive. However, our measures contain a large proportion of zeros, which makes log transformations problematic.

Recent work by Chen and Roth (2024) highlights how applying a log transformation can introduce important biases. Specifically, adding a constant to handle zeros - e.g., log(y + 1) - effectively translates into rescaling the outcome variable in a way that can arbitrarily inflate or deflate estimated effects. Formally, any treatment effect estimate becomes a function of a scaling factor (denoted *"a"* in Chen and Roth (2024)), which depends on both the chosen constant and the distribution of the dependent variable. In the worst case, researchers could manipulate estimated effects simply by adjusting this scaling factor.

There are many contexts in economic research where data might seem suitable for such transformations, creating potential sources of faulty results. For example, if the dependent variable is, e.g., hours worked, changing the unit from hours to days or weeks would improperly affect the coefficient after transformation – a clear violation of sound econometric practice. While such rescaling concerns are less pronounced for count data (like retweets or likes), the broader problem remains: log transformations with zeros can produce misleading or unstable estimates.

In light of these issues, for this study we adopt Poisson Pseudo-Maximum Likelihood (PPML) regression models as our preferred method. Poisson models are particularly suitable for count data. These models naturally handle zeros, avoiding the need for arbitrary transformations that could make results misleading. They also yield coefficients that can be interpreted similarly to semi-elasticities, maintaining a clear and meaningful link between model estimates and real-world effects.

In the remainder of this section, we present some alternative approaches. Although we argue above that Poisson represents the best method for our analysis, we provide these alternative models to offer a comparison and to improve understanding of the main results. In particular, we examine three approaches: OLS without any transformation, OLS applied after log-transforming and asinh-transforming the dependent variables, and the piecewise model proposed by Chen and Roth (2024), which distinguishes between extensive and intensive margin effects.

D.6.1 Linear Probability Models

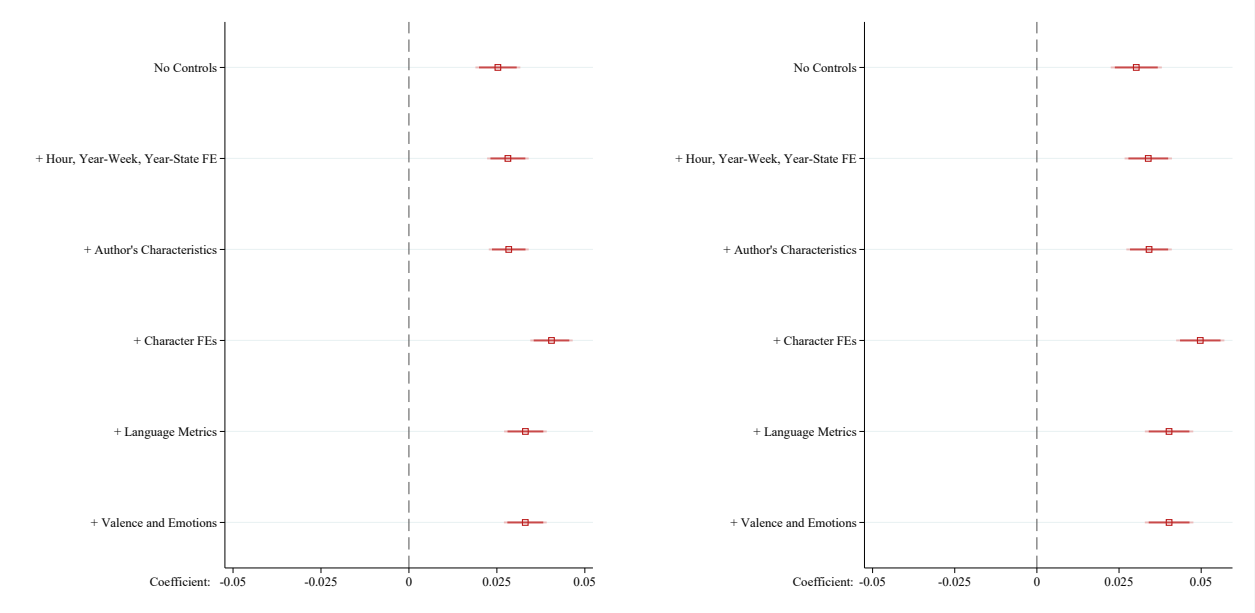
In the first part of this exercise, we reproduce the results using simple linear probability models. We do so leaving the dependent variable unvaried. The particular distribution of retweets and likes, following a power law, is problematic when using OLS and leads to violations of the OLS assumptions regarding the distribution and variance of errors. Nevertheless, we want to provide a comparison with OLS to highlight the differences from what we argue is the most appropriate modeling choice.

In the second part of this exercise, instead of Poisson we follow the other common approach in economics: applying a transformation to the count outcome data and then estimate a linear probability model with OLS. Figure D.7a and Figure D.7b reproduce the main result from Figure 7 using

the most common log(y+1) and inverse hyperbolic sine (asinh(y)) transformations. There are key similarities in the results, but also some differences in the details and sensitivity to controls. First, the key and most important similarity is the robustness of the results. With either transformation, all results are clearly positive and statistically significant at least at the 1% level. It is also still the case that compared to the main specification, adding language metrics as well as valence and emotions leads to a small attenuation in the coefficient. Differences are that there is no consistent pattern of coefficient shrinking or increasing with adding more controls. Initially, more restrictive sets of FE are linked to larger point estimates, but the addition of language metrics, as well as valence and emotions, lowers the coefficient somewhat. Overall, the results are very robust with respect to sign and significance.

Figure D.7: ***Impact of Political Narratives on Virality - Log and Asinh Transformation***

(a) Transformation: $log(Y + 1)$ **(b)** Transformation: Inverse Hyperbolic Sine of Y



Notes: The figure shows the coefficients of OLS regressions testing the effect of featuring at least one character-role, compared to featuring characters only in a neutral role, on virality, measured as the count of retweets. In Figure D.7a the dependent variable is log transformed after adding one unit. In Figure D.7b we transform the dependent variable using the inverse hyperbolic sign transformation. The x-axis reports coefficient estimates. We do not report percentage changes because these depend on the level of the dependent variable. Panels display results from increasingly restrictive models. The first model includes only the indicator variable for containing a character-role and clusters standard errors at the week level. The second model adds hour and year-state fixed effects. The third model accounts for author characteristics (verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status). The fourth model adds character fixed effects. The fifth and sixth models include, respectively, language metrics and valence/emotions.

D.6.2 Extensive and Intensive Margin Model

We also adapt one of the suggestions proposed by Chen and Roth (2024). Specifically, we develop a piecewise regression that combines two models: one capturing the extensive margin effect and the other capturing the intensive margin effect of featuring narratives on virality. The dependent variable is transformed differently for each model. For the first model:

$$y^{*1} = \begin{cases} y = 1 & \text{if } y > 0 \\ y = 0 & \text{if } y = 0 \end{cases}$$

For the second model:

$$y^{*2} = \begin{cases} log(y) & \text{if } y > 0 \\ . & \text{if } y = 0 \end{cases}$$

The first model captures the extensive margin. It includes all observations but transforms any positive value of the dependent variable into 1, while leaving zeros unchanged. This approach captures whether featuring narratives increases the likelihood of any engagement with a tweet. The second model captures the intensive margin. It retains only observations where the dependent variable is greater than zero. For this subset, we apply a log transformation to the dependent variable and estimate the model using OLS. This approach addresses a different question: does featuring narratives increase or decrease the intensity of virality, conditional on receiving some engagement? In this model, the coefficients can be interpreted as semi-elasticities.

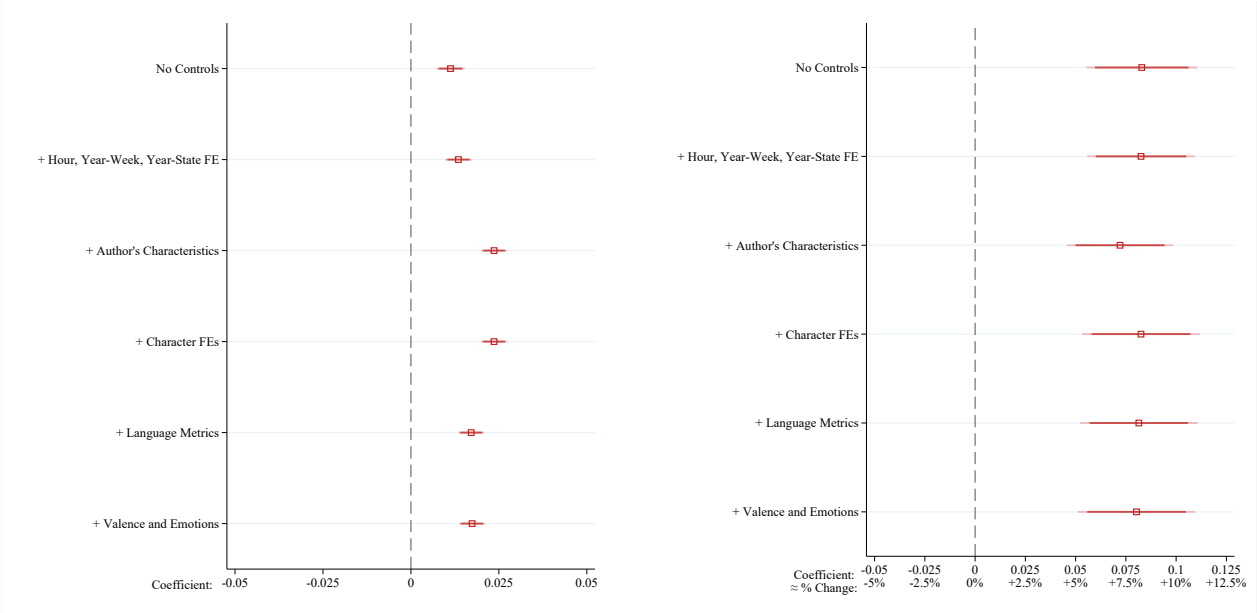
Figure D.8a and Figure D.8b show the extensive and intensive margin effects, respectively. The results show a clear pattern. In both models, political narratives have a positive and statistically significant impact on virality. Featuring a narrative increases both the likelihood of any retweet and the intensity of retweets.

Summary

While the Poisson regression remains the theoretically preferred model for our data, the alternative specifications confirm that our core findings are robust across different modeling choices. At the same time, these exercises highlight the potential pitfalls of common transformations and underscore the value of using models tailored to the data structure and distribution.

Figure D.8: ***Impact of Political Narratives on Virality - Extensive and Intensive Margins***

(a) Extensive Margin **(b)** Intensive Margin



Notes: The figure shows the coefficients of OLS regressions testing the effect of featuring at least one character-role, compared to featuring characters only in a neutral role, on virality, measured as the count of retweets. In Figure D.8a the dependent variable is first transformed to have $Y = 1$ if $Y > 0$, then a Linear Probability Model is applied. In Figure D.8b we drop $Y = 0$ cases and then apply a log transformation. The regression models include relevant tweets, defined as those featuring at least one character from our list. We include only character-roles that appear at least 100 times, thus excluding US Republicans-victim, Emission Pricing-victim, Regulations-victim, and Green Tech-victim. The x-axis reports coefficient estimates, and for Figure D.8b the corresponding percentage change rounded to the closest unit and computed as follows: $\approx \beta * 100$. Panels display results from increasingly restrictive models. The first model includes only the indicator variable for containing a character-role and clusters standard errors at the week level. The second model adds hour and year-state fixed effects. The third model accounts for author characteristics (verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status). The fourth model adds character fixed effects. The fifth and sixth models include, respectively, language metrics and valence/emotions.

D.7 Impact of Narratives on Popularity

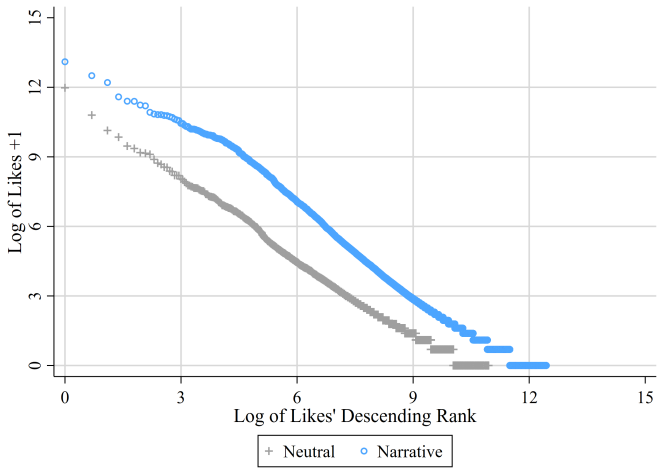
In this section of Appendix D, we check the robustness of our results by examining how narratives affect tweet popularity, measured by the number of likes. We leave these results out of the main paper for two reasons. First, the findings for popularity are very similar to those for virality, which we measure using retweet counts. Second, we believe retweets are a better measure of virality because they capture not just approval and endorsement – as likes might – but also the actual spread of content.

We begin by examining how political narratives influence the distribution of likes. Specifically, we compare tweets that feature a political narrative with those that do not. Figure D.9 shows the Log-Log Rank Distribution of likes, separating tweets that include at least one character-role from those that only present characters in a neutral form. The x-axis reports the logarithm of the

tweet's rank, where rank 1 corresponds to the most liked tweet in the dataset. The y-axis reports the logarithm of the number of likes, plus one.

The figure shows that the distribution of popularity is also highly skewed, similar to the case of retweets. Moreover, at every point along the rank distribution, tweets containing political narratives tend to receive more likes – mirroring the pattern observed for retweets. Compared with the retweet distribution presented in the main paper, the curve for likes appears even steeper. This suggests that the most liked tweets receive more likes than the most retweeted tweets receive retweets, and that the drop-off from the most to the least liked tweets is even steeper. Overall, this descriptive evidence indicates that popularity follows patterns similar to virality and that tweets featuring political narratives consistently attract more likes than neutral ones.

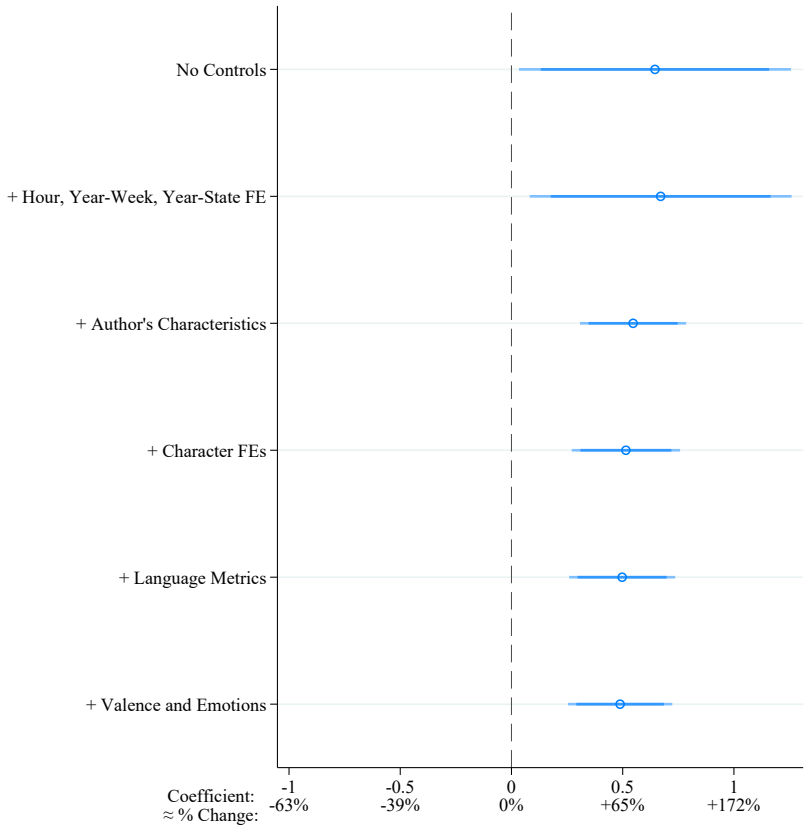
Figure D.9: ***Popularity of Political Narratives in Relevant Tweets (United States, 2010-2021)***



Notes: The figure shows the log–log rank distribution of likes for relevant tweets, distinguishing between those with at least one character–role (blue) and those with characters only in a neutral form (gray). The x-axis plots the logarithm of the rank, with rank 1 corresponding to the most liked tweet in the sample. The y-axis plots the logarithm of like counts (plus one). The slope of the curve indicates how quickly the distribution declines from the most popular to the least popular tweets: a steeper slope means engagement is clustered in a handful of viral tweets, whereas a flatter slope indicates that engagement is more evenly distributed. The vertical position at a given rank reflects relative popularity: a higher curve means tweets at that rank receive more engagement than tweets at the same rank in a dataset with a lower curve. Figure 5 in the paper shows the same for virality (retweets) of tweets.

The regression results for likes in Figure D.10 align closely with those for virality. Consistent with the findings on retweets, political narratives have a clear, positive, and statistically significant effect on popularity. Across all model specifications, tweets that include at least one character-role receive more likes than those with only neutral characters.

Figure D.10: ***Regression Results - Impact of Political Narratives on Popularity (likes)***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regression models testing the effect of featuring at least one character-role vs. featuring characters only in a neutral role on popularity, measured as the count of likes. The x-axis reports coefficient estimates along with the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. Panels display results from increasingly restrictive models. The first model includes only the indicator variable for containing a character-role and clusters standard errors at the week level. The second model adds hour, week of the year, and year-state fixed effects. The third model controls for author characteristics: verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status. The fourth model adds character fixed effects. The fifth and sixth models include, respectively, language metrics and valence/emotions. Figure 7 in the paper shows the same for virality (retweets) of tweets.

D.8 Impact of Political Narratives on Conversation

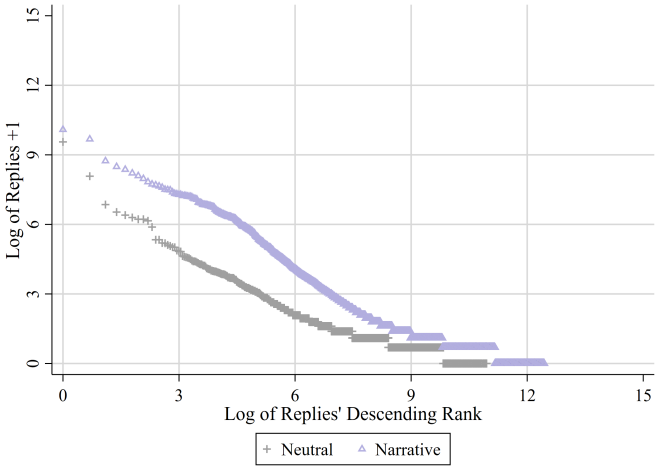
In this section of Appendix D.8, we provide a robustness check on the impact of political narratives by examining their effect on conversation. We measure conversation using the count of comments that a post receives.

We begin by exploring the distribution of comments in tweets featuring a political narrative and in tweets featuring characters only in neutral framing. Similar to what is presented in Subsection 4.3 of the paper for virality, Figure D.11 shows the Log-Log Rank Distribution of the count of comments, in tweets featuring a narrative (triangle shape) and tweets featuring no narrative (cross shape). The x-axis represents the logarithm of the rank of tweets based on their comment count, with rank 1

corresponding to the tweet that received the most comments. The y-axis represents the logarithm of the number of comments plus one.

The distribution reveals several interesting insights. First, the distribution of comments appears to be highly skewed, consistent with what we observe for retweets. The linear shape of the log-log rank distribution suggests that the data generation process follows an exponential structure, with most tweets receiving few comments and a small number receiving many. Second, compared to retweets, the curves are flatter and shifted downward, reflecting the smaller number of comments overall. Finally, the figure provides an initial indication of the effect of narratives. The curve for narrative tweets lies above that for neutral tweets. At each rank, tweets featuring a narrative receive more comments, suggesting that political narratives may not only spark more engagement but also encourage greater participation through conversation.

Figure D.11: ***Conversation on Political Narratives in Relevant Tweets (United States, 2010-2021)***



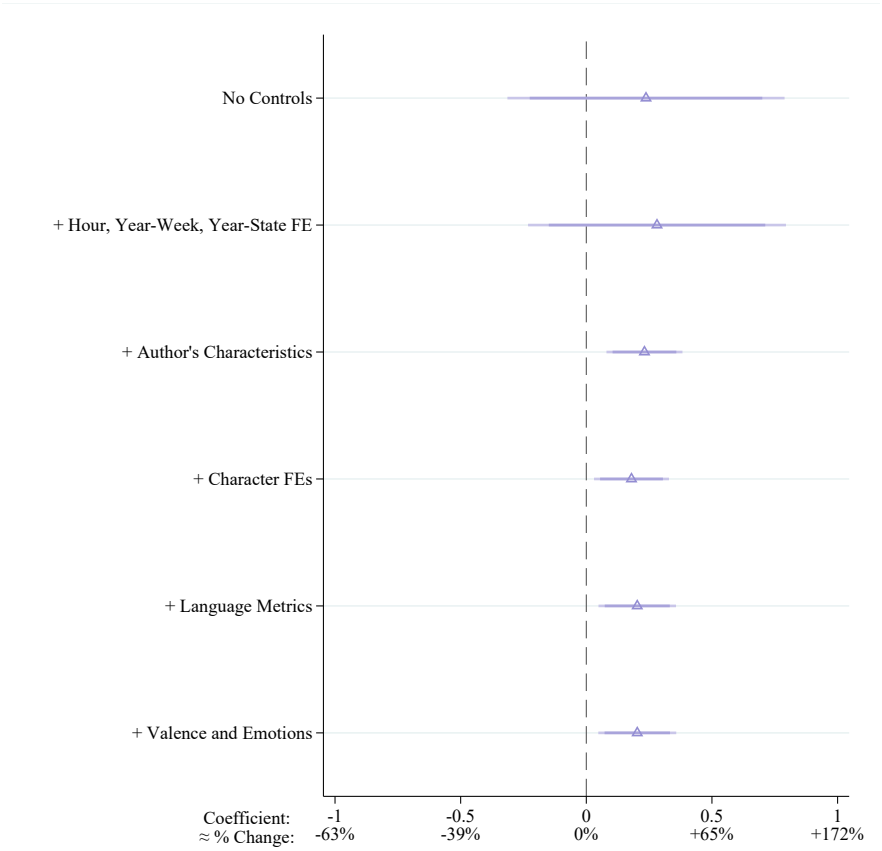
Notes: The figure shows the log–log rank distribution of replies for relevant tweets, distinguishing between those with at least one character–role (blue) and those with characters only in a neutral form (gray). The x-axis plots the logarithm of the rank, with rank 1 corresponding to the tweet with the most replies in the sample. The y-axis plots the logarithm of reply count (plus one). The slope of the curve indicates how quickly the distribution declines from the most popular to the least popular tweets: a steeper slope means engagement is clustered in a handful of viral tweets, whereas a flatter slope indicates that engagement is more evenly distributed. The vertical position at a given rank reflects relative popularity: a higher curve means tweets at that rank receive more engagement than tweets at the same rank in a dataset with a lower curve. Figure 5 in the paper shows the same for virality (retweets) of tweets.

We move from descriptive to regression analysis by reproducing the results of Figure 7 for conversation. Figure D.12 presents a coefficient plot showing the results of models that mirror those used for virality as a robustness check. The models are estimated using Poisson Pseudo-Maximum Likelihood and employ increasingly restrictive specifications, moving from the top panel to the bottom panel. The additions to each specification are indicated in the panel titles. Coefficients can be interpreted as percentage changes using the following transformation: $\approx e^{\beta} - 1$.

The results are mostly consistent with those found for virality in the paper, and popularity in the appendix. Although the first two models have large confidence intervals – suggesting noisier

measures compared to retweets and likes – the overall effect indicates that narratives spark more conversation than tweets in which characters appear only in neutral framing. The author fixed effects play an important role by reducing excess noise and stabilizing the coefficients around 0.2, which corresponds to roughly a 22% increase in the number of comments on tweets featuring Political Narratives. The size of this coefficient is about half that observed for retweets and likes, but narratives still appear to play an important role in driving conversation.

Figure D.12: ***Regression Results - Impact of Political Narratives on Conversation***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regression models testing the effect of featuring at least one character-role vs. featuring characters only in a neutral role on conversation, measured as the count of replies. The x-axis reports coefficient estimates along with the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. Panels display results from increasingly restrictive models. The first model includes only the indicator variable for containing a character-role and clusters standard errors at the week level. The second model adds hour, week of the year, and year-state fixed effects. The third model controls for author characteristics: verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status. The fourth model adds character fixed effects. The fifth and sixth models include, respectively, language metrics and valence/emotions. Figure 7 in the paper shows the same for virality (retweets) of tweets.

D.9 Additional Details on Regression Models

In this section, we complement the main results presented in Section 5. First, we report Table D.1, which lists the coefficients displayed in Figure 7. The table displays results from increasingly restrictive models. Along with the coefficients, we also report the percentage changes.

Table D.1: ***Regression Results – Impact of Character-Role Narratives on Virality***

Dependent Variable	**Virality: Retweets (PPML)**					
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Political Narrative	0.647	0.707	0.603	0.501	0.468	0.464
	(0.243)	(0.238)	(0.131)	(0.137)	(0.137)	(0.133)
	[0.008]	[0.003]	[0.000]	[0.000]	[0.001]	[0.000]
Hour FE		✓	✓	✓	✓	✓
Year-Week FE		✓	✓	✓	✓	✓
Year-State FE		✓	✓	✓	✓	✓
Author Characteristics			✓	✓	✓	✓
Character FE				✓	✓	✓
Language Metrics					✓	✓
Valence and Emotions						✓
Percentage Change	91%	102.8%	82.7%	65%	59.7%	59%
Mean Outcome (Control Group)	2.15	2.15	2.15	2.15	2.15	2.15
Observations	309,744	309,531	309,531	309,531	309,531	309,531

Notes: The table displays the coefficients of Poisson Pseudo-Maximum Likelihood regression models testing the effect of featuring at least one character-role vs. featuring characters only in a neutral role on virality, measured as the count of retweets. Columns display results from increasingly restrictive models. The first model includes only the indicator variable for containing a character-role and clusters standard errors at the week level. The second model adds hour, week of the year, and year-state fixed effects. The third model controls for author characteristics: verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status. The fourth model adds character fixed effects. The fifth and sixth models include, respectively, language metrics and valence/emotions. Plot of reference is Figure 7.

Moreover, we complement this section with Table D.2, an alternative version of Figure 9a in which we computed the coefficients using OLS instead of Poisson Pseudo-Maximum Likelihood regressions.

Table D.2: ***Regression Results – Heterogeneous Impact of Narratives on Virality (OLS Model)***

Dependent Variable	**Retweets Count**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Human	1.193	1.127		
	(0.705)	(1.090)		
	[0.091]	[0.301]		
Instrument	1.161	1.103		
	(0.652)	(0.489)		
	[0.075]	[0.025]		
Human × Instrument		0.133		
		(1.357)		
		[0.922]		
Sum of CRs			0.862	1.751
			(0.361)	(0.655)
			[0.017]	[0.008]
Sum of CRs Squared				-0.285
				(0.138)
				[0.039]
Mean Outcome Reference Group	2.151	2.151	2.151	2.151
Adj. R2	0.10	0.10	0.10	0.10
Obs.	309743	309743	309743	309743

Notes: The table reports OLS estimates for heterogeneous effects of narratives on virality (retweet counts). Columns (1)–(2) isolate heterogeneity by character type. Column (1) includes dummy variables for tweets featuring human and instrument characters, with the omitted category being tweets containing only neutral characters. Column (2) augments this specification by including an interaction term identifying tweets that feature both human and instrument characters. Columns (3)–(4) explore the cumulative effect of featuring additional character-roles in a tweet: Column (3) includes the total count of character-roles as a regressor, while Column (4) adds the squared term to test for non-linear effects. All regressions control for author characteristics (verified status, number of followers/followings, total tweets created, party affiliation, religiosity, higher education, and parenthood) and include character fixed effects. We also include hour, week-of-year, and year–state fixed effects. Standard errors are clustered at the week level, covering the full time frame. Paper Figure 9a is the reference table.

D.10 Sensitivity to Unobservables

In this section, we assess the robustness of the estimated effect of political narratives on virality to potential omitted-variable bias. To do so, we estimate the same baseline specification under two alternative transformations of the dependent variable, the log transformation, log(Y+1), and the inverse hyperbolic sine, asinh(Y), and compute the corresponding Altonji–Elder–Taber (AET) ratios (Altonji, Elder, and Taber 2005). We choose these transformations to align with Subsection D.6.

These transformations ensure that the results are not driven by the particular functional form of the outcome and provide a stable basis for evaluating sensitivity to unobservables.

Across both transformations, the estimated AET ratios remain large and stable. In the most saturated specification, the AET ratio equals 4.44 for the log outcome and 4.24 for the asinh outcome. These values imply that unobserved factors would need to be more than four times as strongly correlated with both the treatment and the outcome as the observed controls in order to fully explain away the estimated effect. The similarity of these ratios across transformations reinforces the conclusion that the relationship between political narratives and tweet virality is highly robust to omitted-variable bias: only implausibly strong selection on unobservables could eliminate the observed effect.

Table D.3: ***Robustness Check – Linear Models and Selection Sensitivity***

Dependent Variable:	*log(retweets + 1)*		*asinh(retweets)*	
Regression Model	**SHORT**	**FULL**	**SHORT**	**FULL**
	(1)	(2)	(3)	(4)
	log(Y + 1)	*log(Y + 1)*	*asinh(Y)*	*asinh(Y)*
Political Narrative	0.041	0.033	0.050	0.040
	(0.003)	(0.003)	(0.004)	(0.004)
Hour FE	✓	✓	✓	✓
Year-Week FE	✓	✓	✓	✓
Year-State FE	✓	✓	✓	✓
Author Characteristics		✓		✓
Character FE		✓		✓
Language Metrics		✓		✓
Emotions and Valence		✓		✓
Adj. R^2	0.167	0.174	0.164	0.172
Observations	309,743	309,743	309,743	309,743
AET Ratio	.	4.44	.	4.24

Notes: The dependent variables for Columns (1) and (2) are log-transformed retweet counts. In Columns (3) and (4) we take the inverse hyperbolic sine transformations of retweet counts. The SHORT models include fixed effects for hour, week, and year-state, and controls for author characteristics and character fixed effects. The FULL models additionally controls for language metrics and for valence and emotions. The AET ratio is from (Altonji, Elder, and Taber 2005)

E Additional Output: Experimental Data

E.1 Descriptive statistics for experiments

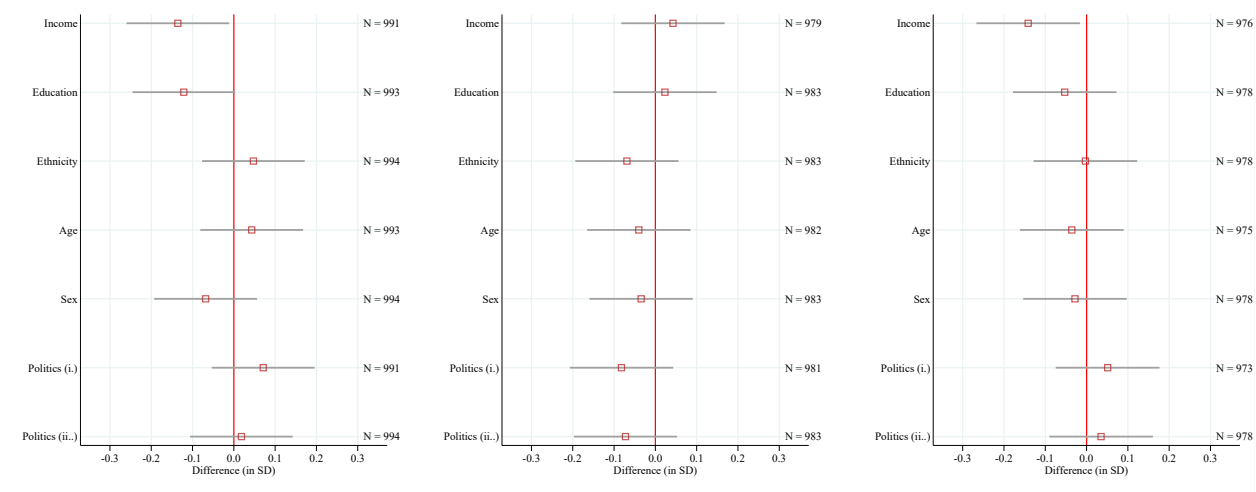
In this section of Appendix E, we provide additional information on our online pre-registered experiments. We conducted three separate studies, each with a representative sample of the U.S. population based on age, ethnicity, gender, and political affiliation. Within each experiment, participants were randomly assigned to either a control group or a treatment group. Both groups were told that the survey aimed to study the impact of exposure to social media and were shown a fictitious feed of three posts resembling Twitter or Facebook. Two posts were identical across groups and served as obfuscation. The third post differed: in the control group it conveyed a piece of factual information and depicted characters in a neutral way, while in the treatment group it presented the same information but framed the characters within a drama-triangle narrative.

In each experiment, the control–treatment pair of tweets featured different characters, except for Green Tech, which appeared in all three designs. We label the experiments according to the characters included in the narrative treatment. The first is the Hero–Hero experiment, featuring Green Tech-Hero and US People-Hero. The second is the Hero–Hero–Villain experiment, featuring Green Tech-Hero, Regulations-Hero, and Fossil Industry-Villain. The third is the Villain–Villain–Hero experiment, featuring Green Tech-Hero, Corporations-Villain, and Fossil Industry-Villain. The day after exposure, participants were invited to complete a short follow-up survey, identical across experiments, to assess recall and other outcomes. In the main sections of the paper, we provide further details on design choices; below, we also provide links to all experimental questionnaires for transparency.

- Main Experiment - Hero-hero: Click here
- Main Experiment - Hero-Hero-Villain: Click here
- Main Experiment - Villain-Villain-Hero: Click here
- Follow-up Experiment (all conditions): Click here

At the core of our design is the assumption that randomization was successful, ensuring that the control and treatment groups in all three experiments are comparable on average. Figure E.1 presents balance tests for a range of participant characteristics. We obtain the coefficients by regressing each individual feature on an indicator of assignment to the treatment group. With only two exceptions, all tests confirm balance between treatment and control groups. Income appears slightly lower in the treatment group for both the Hero–Hero experiment and the Villain–Villain–Hero experiment. In principle, our main analyses already control for this set of personal characteristics, but for completeness, we also report results without controls later in this appendix.

Figure E.1: ***Balance of Individual Characteristics by Experiment Wave***



(a) Hero-Hero **(b)** Hero-Hero-Villain **(c)** Villain-Villain-Hero

Notes: The figure shows control-treatment balance tests for several individual characteristics among the participants of each experiment. Panel (a) shows results for the experiment Hero-Hero, Panel (b) shows result for the experiment Hero-Hero-Villain, and Panel (c) for Villain-Villain-Hero. Tests are performed by regressing each individual characteristic on a dummy variable that takes value 1 if the participant is in the treatment group. We use robust standard errors.

E.2 Additional Details on Experimental Output

In this section we provide additional details on the experimental output that we describe in the paper. In particular we provide the underlying models for each coefficient plot in Section 6. Table E.1 provides the models for the Paper Figure 10a, Table E.2 for the Paper Figure 10b, Table E.3 about the donation results in Figure 11. Table E.4 and Table E.5 provide additional details on our memory results, from the Paper Figure 12b and Figure 12c.

Table E.1: ***Experimental Results - The Impact of Political Narratives on Beliefs***

Dependent Variable	**Expectation for the Future:**					
	Forecast	**Confidence**	**Forecast**	**Confidence**	**Forecast**	**Confidence**
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	2.487	3.426	0.831	0.331	-5.022	-0.358
	(1.368)	(1.680)	(1.425)	(1.740)	(1.387)	(1.803)
	[0.069]	[0.042]	[0.560]	[0.849]	[0.000]	[0.843]
Controls	✓	✓	✓	✓	✓	✓
Experiment: Hero-Hero	✓	✓				
Experiment: Hero-Hero-Villain			✓	✓		
Experiment: Villain-Villain-Hero					✓	✓
Mean Outcome Control Group	39.02	46.02	43.67	48.54	42.32	48.15
Observations	987	987	976	976	968	968

Notes: The table displays the coefficients of OLS regression models providing insights on two outcomes: first, the participants' forecast, as answer to the question *'What percentage of US energy do you predict will come from renewable sources and green technology by the year 2035? Indicate a number between 0 and 100.'* (in Columns 1, 3, 5), second, their confidence in the forecast, as answer to *'Your response on the previous screen suggests that by 2035, [x]% of US energy will come from renewable sources and green technology. How certain are you that the actual share of renewable energy in 2035 will be between [x-5] and [x+5]%?'* (in Columns 2, 4, 6). Columns 1 and 2 show results for the Hero-Hero experiment, columns 3 and 4 for the Hero-Hero-Villain experiment, and 5 and 6 for the Villain-Villain-Hero experiment. All models include income, education, political preference, age, and sex as controls. We use robust standard errors. Figure 10a in the paper is the reference plot.

Table E.2: ***Experimental Results - The Impact of Political Narratives on Stated Preferences***

Dependent Variable	**Stated Preferences: Policy Support**		
	Coeff./SE/p-value		
	(1)	(2)	(3)
Treatment	0.061	0.054	0.094
	(0.090)	(0.086)	(0.117)
	[0.495]	[0.533]	[0.421]
Controls	✓	✓	✓
Experiment: Hero-Hero	✓		
Experiment: Hero-Hero-Villain		✓	
Experiment: Villain-Villain-Hero			✓
Mean Outcome Control Group	5.70	5.72	4.19
Observations	987	976	968

Notes: The table displays the coefficients of OLS regression models providing insights on the support or opposition for a policy or law that is in line with the content of the narrative tweet in our experiments. In Column 1, we show results for the Hero-Hero experiment, where participants were asked whether they would support a policy that reduces the cost of residential renewable systems. In Column 2, we show results for the Hero-Hero-Villain experiment, where participants were asked whether they would support increasing transparency and accountability for energy companies. In column 3, we show results for the Villain-Villain-Hero experiment, where people were asked whether they would support raising taxes on fossil fuels. All models include income, education, political preference, age, and sex as controls. We use robust standard errors. Figure 10b in the paper is the reference plot.

Table E.3: ***Experimental Results - Impact of Political Narratives on Revealed Preferences***

Dependent Variable	**Revealed Preference: Incentivized Donation**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Treatment	0.734	0.530	0.584	0.562
	(0.497)	(0.469)	(0.480)	(0.272)
	[0.140]	[0.259]	[0.224]	[0.039]
Controls	✓	✓	✓	✓
Experiment: Hero-Hero	✓			✓
Experiment: Hero-Hero-Villain		✓		✓
Experiment: Villain-Villain-Hero			✓	✓
Experiment FE				✓
Mean Outcome Control Group	6.52	6.40	6.73	6.55
Observations	987	976	968	2,931

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' revealed preferences. We measure revealed preferences with the decision to donate to an association promoting sustainable development and local/national projects to support Green Tech diffusion. The decision is incentivized with a lottery: participants could be selected to win 25$ and had to allocate the amount between themselves and the association. No restrictions on the allocation were given. Columns 1, 2, and 3 show results for the single experiments, Column 4 shows results for the pooling together all experiments, and it includes experiment fixed effects. All models include income, education, political preference, age, and sex as controls. We use robust standard errors. Figure 11 in the paper is the reference plot.

Table E.4: ***Experimental Results - The Impact of Political Narratives on Memory (Pooled Sample)***

Dependent Variable	**Information Retention:**			
	Facts		**Character**	
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Treatment	0.001	-0.002	0.061	0.049
	(0.012)	(0.013)	(0.016)	(0.021)
	[0.957]	[0.901]	[0.000]	[0.018]
Controls	✓	✓	✓	✓
Experiment FEs	✓	✓	✓	✓
Day of the Experiment	✓		✓	
Day After the Experiment		✓		✓
Mean Outcome Control Group	0.13	0.10	0.69	0.45
Observations	2,931	2,280	2,931	2,269

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' memory. Columns 1 and 2 show the effect on information retention, respectively in the day of the main experiment and a day later. Participants were asked to remember the factual information reported in the tweet. Column 3 and 4 show the effect on recalling the characters present in the text. Columns 5 and 6 show the effect on recalling the characters framed in their role. The dependent variables for columns from 3 to 6 are obtained encoding an open-ended question that asked participants to recall anything from the text they saw. All models include income, education, political preference, age, and sex as controls, and experiment FEs. We use robust standard errors. Paper Figure 12a and Figure 12b are the reference plots.

Table E.5: ***Experimental Results - The Impact of Political Narratives on Memory of Roles***

Dependent Variable	**Information Retention:**				
	Hero	**Hero**	**Villain**	**Hero**	**Villain**
	Coeff./SE/p-value				
	(1)	(2)	(3)	(4)	(5)
Treatment	0.169	-0.106	0.133	-0.139	0.636
	(0.075)	(0.075)	(0.059)	(0.064)	(0.077)
	[0.025]	[0.156]	[0.025]	[0.032]	[0.000]
Controls	✓	✓	✓	✓	✓
Experiment: Hero-Hero	✓				
Experiment: Hero-Hero-Villain		✓	✓		
Experiment: Villain-Villain-Hero				✓	✓
Mean Outcome Control Group	1.35	1.25	0.67	1.01	0.81
Observations	784	792	792	693	693

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' memory of characters divided by role. Column 1 includes only the Hero–Hero experiment and tests recall of Green Tech and US People (both heroes). Columns 2 and 3 include the Hero–Hero–Villain experiment, showing effects on recall of Green Tech and Regulations (heroes, column 2) and Fossil Industry (villain, column 3). Columns 4 and 5 cover the Villain–Villain–Hero experiment, showing effects on recall of Green Tech (hero, column 4) and Fossil Industry/Corporations (villains, column 5). All models include income, education, political preference, age, and sex as controls. We use robust standard errors. Figure 12c in the paper is the reference plot.

E.3 The Impact of Valence and Anger

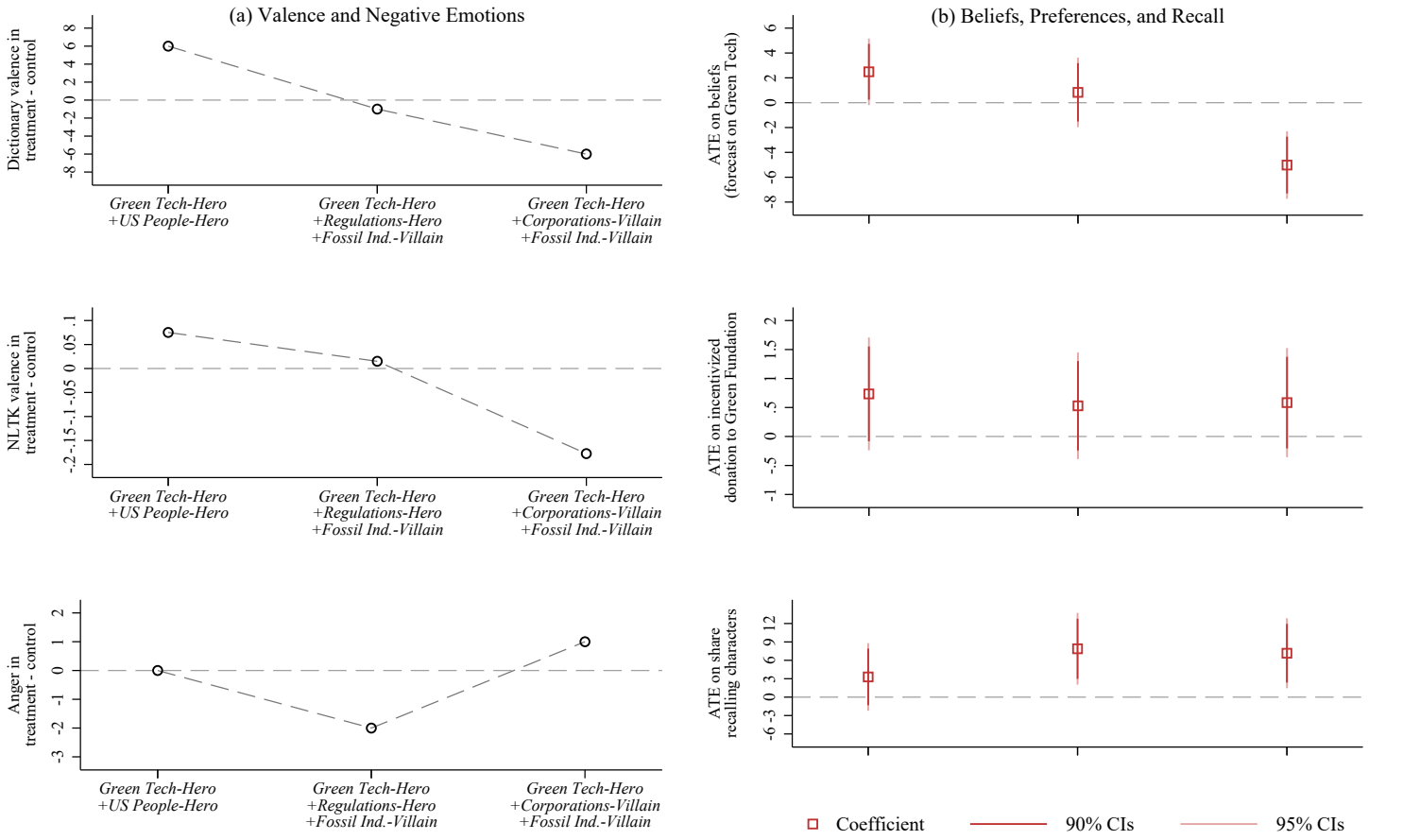
One potential concern with our experiments is that the results described in the main paper might be driven primarily by emotions rather than by the use of political narratives. The marketing literature highlights the central role of emotions and arousal in determining the virality of content (Berger 2016; Berger 2011). For example, Berger (2011) argues that much of a message's contagiousness can be explained by whether it features emotions such as disgust, which heighten the receiver's arousal. Guided by this evidence, we explore the correlation between emotions and our experimental results.

While we recognize that emotions are an important ingredient in building and communicating political narratives, our analysis throughout the paper shows that they are not sufficient to explain their effect. The roles of villain, hero, and victim capture deeper mental constructs that extend beyond emotional tone. To test this directly, recall that participants in our experiments were shown a social media feed containing tweets about climate change. In the control group, a tweet conveyed factual information with characters portrayed neutrally; in the treatment group, the same information was presented but with the characters framed in drama triangle roles. If emotions alone captured the essence of political narratives, the emotional content of the tweets should fully explain the differences in outcomes we observe. Building on this idea, we test whether differences in valence – defined as the balance of positive emotions minus negative ones – between treatment and control tweets account for the direction and significance of our results.

Figure E.2 sheds light on this issue. The left side of the figure reports the differences in valence and anger between the treatment and control tweets in each experiment. In the top panel, we compute valence as the number of positive emotion words (joy, trust, anticipation, and surprise) minus the number of negative emotion words (anger, disgust, fear, and sadness), following the NRC dictionary (Mohammad and Turney 2013). We then calculate, for each experiment, the difference between the treatment tweet's valence and that of the control tweet. A positive value indicates that the treatment tweet carried a more positive tone, whereas a negative value signals greater negativity. The middle panel repeats this exercise using the NLTK sentiment package as an alternative measure of valence. Finally, the bottom panel focuses specifically on anger, showing the difference in the frequency of anger-related words between treatment and control tweets. The right side of Figure E.2 provides a condensed summary of our experimental results. The top panel shows the effect of the narrative treatment on beliefs, measured by participants' forecasts of the future share of energy generated from green technologies. The middle panel presents the impact on revealed preferences, captured through donations to green technology institutions. The bottom panel reports our main results on memory, measured as the likelihood of recalling characters from the narratives. For a detailed discussion of these outcomes, we refer back to the main body of the paper.

Despite being purely correlational and descriptive, the figure delivers a clear message: there is no consistent link between the emotional tone of the treatment tweets and the experimental outcomes. For beliefs, the relationship with net valence appears plausible: tweets with a more negative tone seem to elicit a more pessimistic outlook on the future. However, for both donations and memory, the narrative treatment has a positive effect regardless of whether the tweet's valence is positive or negative. Finally, when focusing specifically on anger – the emotion most likely to heighten arousal – there is virtually no correlation with the experimental results. These findings should be interpreted with caution, but they suggest an important takeaway: while emotions, valence, and arousal certainly play a role, they cannot by themselves explain the effects we observe. Narratives structured around hero, villain, and victim roles capture a deeper mechanism of influence, one that goes beyond emotional tone alone.

Figure E.2: ***Emotional Content in the Experimental Design***



Notes: Panel a of Figure E.2 displays the difference in the corresponding emotion score between the treatment tweet and the control tweet, for each of the three experimental designs (Hero-Hero, Hero-Hero-Villain, Hero-Villain-Villain). The emotion scores used are valence and anger. Valence is measured either as the difference in the number of positive words and negative words (top) or as a score using the NLTK Python package (middle). Anger is computed as a score using the NLTK Python package (bottom). Emotion and valence use the NRC Emotion Lexicon (Mohammad and Turney (2013)). Each graph of panel (a) represents the difference in the score between the treatment tweet and the control tweet. Figure E.2 panel (b) replicates the main results of the paper. Reference figures are Figure 10a, Figure 11, Figure 12c.

F Robustness Checks: Experimental Data

F.1 Recall of the Factual Information

One important insight from our experimental results concerns the nature of memory recall. On the one hand, participants in the treatment groups remembered the characters featured in the tweets much more – both on the day of the experiment and the day after. On the other hand, we find no significant difference between the treatment and control groups in the recall of factual information. In the paper, factual recall is measured using strict encoding: participants were asked to reproduce exactly the factual information reported in the tweets. A potential concern is that the absence of differences could simply reflect the limited scope of this measure. To address this, we provide a robustness test in this section. Even when we relax the encoding and accept answers within broader intervals around the correct value, no detectable treatment–control difference emerges.

Table F.1 and Table F.2 report robustness checks on the recall of factual information, measured on the day of the experiment and the day after, respectively. In both tables, Column 1 reproduces the baseline specification from the paper, while Columns 2, 3, and 4 progressively relax the coding of correct answers by accepting responses within ±10, ±20, and ±50 units of the true value. The results are clear: across all specifications, there is no statistically significant difference in factual recall between treatment and control groups.

Table F.1: ***Experimental Results - Different Encoding of Factual Recall on the Experiment Day (Pooled)***

Dependent Variable	**Interval of Recall: 163 +|-**			
	0	**10**	**20**	**50**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Treatment	0.001	0.003	-0.005	-0.012
	(0.012)	(0.015)	(0.016)	(0.018)
	[0.957]	[0.845]	[0.777]	[0.499]
Mean Outcome Control Group	0.12	0.20	0.23	0.35
Observations	2,931	2,931	2,931	2,931

Notes: The table reports OLS regression results on the effect of narratives on participants' memory of the factual information embedded in both control and treatment tweets, measured on the day of the experiment. The dependent variable comes from the question: "How many billion kWhs were generated using solar energy in 2023 in the US? Please indicate your best guess." Column 1 codes the outcome as 1 if the answer is exactly correct; Column 2 as 1 if the answer falls within a 10-unit range; Column 3 within 20 units; and Column 4 within 50 units. All models include observations from all three experiments and experiment fixed effects. Additionally, all models control for income, education, political preference, age, and sex, and use robust standard errors. Figure 12a in the paper is the reference plot.

Table F.2: ***Experimental Results - Different Encoding of Factual Recall on the Follow-Up Day (Pooled)***

Dependent Variable	**Interval of Recall: 163 +|-**			
	0	**10**	**20**	**50**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Treatment	-0.002	0.005	-0.006	-0.016
	(0.013)	(0.016)	(0.018)	(0.020)
	[0.901]	[0.756]	[0.748]	[0.436]
Mean Outcome Control Group	0.11	0.19	0.22	0.34
Observations	2,280	2,280	2,280	2,280

Notes: The table reports OLS regression results on the effect of narratives on participants' memory of the factual information embedded in both control and treatment tweets, measured on the day after of the experiment. The dependent variable comes from the question: "How many billion kWhs were generated using solar energy in 2023 in the US? Please indicate your best guess." Column 1 codes the outcome as 1 if the answer is exactly correct; Column 2 as 1 if the answer falls within a 10-unit range; Column 3 within 20 units; and Column 4 within 50 units. All models include observations from all three experiments and experiment fixed effects. Additionally, all models control for income, education, political preference, age, and sex, and use robust standard errors. Figure 12a in the paper is the reference plot.

F.2 Output Excluding Controls

In this section we present robustness tests for the experimental results. In particular we reproduce all the results of the paper, exlcuding individual characteristics as controls in the models.

Table F.3: ***Manipulation Check - Effectiveness of the Political Narrative Treatment - Excluding Controls***

Dependent Variable	**Mention of Character-Role**							
	Coeff./SE/p-value							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.296	0.276	0.237	0.270	0.359	0.330	0.254	0.315
	(0.030)	(0.031)	(0.031)	(0.018)	(0.033)	(0.034)	(0.031)	(0.019)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Outcome | Recall 1+ Characters					✓	✓	✓	✓
Experiment: Hero-Hero	✓			✓	✓			✓
Experiment: Hero-Hero-Villain		✓		✓		✓		✓
Experiment: Villain-Villain-Hero			✓	✓			✓	✓
Experiment FE				✓				✓
Mean Outcome Control Group	0.33	0.34	0.44	0.37	0.45	0.52	0.64	0.54
Observations	994	983	978	2,955	744	690	702	2,136

Notes: The table reports OLS estimates from manipulation checks of the narrative treatment. The dependent variable is a binary indicator equal to one if the participant recalled the role assigned to a character in the treatment tweet (e.g., Green Tech as hero in the Hero–Hero experiment, Fossil Industry as villain in the Hero–Hero–Villain experiment). Columns 1–4 report effects on the unconditional likelihood of recalling the character-role, while Columns 5–8 restrict the sample to participants who recalled at least one character from the tweet. Only the treatment group was exposed to characters explicitly framed in roles; control participants saw the same characters presented neutrally. Non-zero recall in the control group therefore reflects participants' prior beliefs or participants' interpretation of the control tweet, while the treatment effect captures the additional role attribution induced by framing. We use robust standard errors. Table G.16 in the paper is the reference table.

Table F.4: ***Experimental Results - The Impact of Political Narratives on Beliefs - Excluding Controls***

Dependent Variable	**Expectation for the Future:**					
	Forecast	**Confidence**	**Forecast**	**Confidence**	**Forecast**	**Confidence**
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	1.838	3.018	0.755	0.037	-4.995	-0.856
	(1.321)	(1.694)	(1.360)	(1.716)	(1.311)	(1.717)
	[0.165]	[0.075]	[0.579]	[0.983]	[0.000]	[0.618]
Experiment: Hero-Hero	✓	✓				
Experiment: Hero-Hero-Villain			✓	✓		
Experiment: Villain-Villain-Hero					✓	✓
Mean Outcome Control Group	39.02	46.02	43.67	48.54	42.32	48.15
Observations	994	994	983	983	978	978

Notes: The table displays the coefficients of OLS regression models providing insights on two outcomes: the participants' forecast, as answer to the question *'What percentage of US energy do you predict will come from renewable sources and green technology by the year 2035? Indicate a number between 0 and 100.'* (in Columns 1, 3, 5), their confidence in the forecast, as answer to *'Your response on the previous screen suggests that by 2035, [x]% of US energy will come from renewable sources and green technology. How certain are you that the actual share of renewable energy in 2035 will be between [x-5] and [x+5]%?'* (in Columns 2, 4, 6). Columns 1 and 2 show results for the Hero-Hero experiment, columns 3 and 4 for the Hero-Hero-Villain experiment, and 5 and 6 for the Villain-Villain-Hero experiment. The models do not include personal characteristics as controls. We use robust standard errors. Figure 10a in the paper is the reference plot.

Table F.5: ***Experimental Results - The Impact of Political Narratives on Stated Preferences - Excluding Controls***

Dependent Variable	**Stated Preferences: Policy Support**		
	Coeff./SE/p-value		
	(1)	(2)	(3)
Treatment	0.040	0.052	-0.011
	(0.092)	(0.092)	(0.126)
	[0.664]	[0.568]	[0.930]
Experiment: Hero-Hero	✓		
Experiment: Hero-Hero-Villain		✓	
Experiment: Villain-Villain-Hero			✓
Mean Outcome Control Group	5.70	5.72	4.19
Observations	994	983	978

Notes: The table displays the coefficients of OLS regression models providing insights on the support or opposition for a policy or law that is in line with the content of the narrative. In Column 1, we show results for the Hero-Hero experiment, where participants were asked whether they would support a policy that reduces the cost of residential renewable systems. In Column 2 we show results for the Hero-Hero-Villain experiment, where participants were asked whether they would support increasing transparency and accountability for energy companies. In Column 3, we show results for the Villain-Villain-Hero experiment, where people were asked whether they would support raising taxes on fossil fuels. The models do not include personal characteristics as controls. We use robust standard errors. Figure 10b in the paper is the reference plot.

Table F.6: ***Experimental Results - Impact of Political Narratives on Revealed Preferences - Excluding Controls***

Dependent Variable	**Revealed Preference: Incentivized Donation**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Treatment	0.540	0.570	0.311	0.474
	(0.477)	(0.458)	(0.469)	(0.270)
	[0.257]	[0.214]	[0.507]	[0.079]
Experiment: Hero-Hero	✓			✓
Experiment: Hero-Hero-Villain		✓		✓
Experiment: Villain-Villain-Hero			✓	✓
Experiment FE				✓
Mean Outcome Control Group	6.52	6.40	6.73	6.55
Observations	994	983	978	2,955

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' revealed preferences. We measure revealed preferences with the decision to donate to an association promoting sustainable development and local/national projects to support Green tech diffusion. The decision is incentivized with a lottery: participants could be selected to win 25$ and had to allocate the amount between themselves and the association. No restrictions on the allocation were given. Columns 1, 2, and 3 show results for the single experiments, Column 4 shows results for the pooling together all experiments, and it includes experiment fixed effects. The models do not include personal characteristics as controls. We use robust standard errors. Figure 11 in the paper is the reference plot.

Table F.7: ***Experimental Results - The Impact of Political Narratives on Memory (Pooled Sample) - Excluding Controls***

Dependent Variable	**Information Retention:**			
	Facts		**Character**	
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Treatment	0.001	-0.002	0.059	0.053
	(0.012)	(0.012)	(0.016)	(0.021)
	[0.917]	[0.847]	[0.000]	[0.011]
Experiment FEs	✓	✓	✓	✓
Day of the Experiment	✓		✓	
Day After the Experiment		✓		✓
Mean Outcome Control Group	0.13	0.13	0.69	0.45
Observations	2,955	2,297	2,955	2,286

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' memory. Columns 1 and 2 show the effect on information retention, respectively in the day of the main experiment and a day later. Participants were asked to remember the factual information reported in the tweet. Column 3 and 4 show the effect on recalling the characters present in the text, respectively in the day of the main experiment and a day later. All models include experiment FEs. We use robust standard errors. Paper Figure 12a and Figure 12b are the reference plots.

Table F.8: ***Experimental Results - The Impact of Political Narratives on Memory of Roles - Excluding Controls***

Dependent Variable	**Information Retention:**				
	Hero	**Hero**	**Villain**	**Hero**	**Villain**
	Coeff./SE/p-value				
	(1)	(2)	(3)	(4)	(5)
Treatment	0.194	-0.126	0.092	-0.157	0.620
	(0.073)	(0.072)	(0.058)	(0.062)	(0.077)
	[0.008]	[0.082]	[0.109]	[0.012]	[0.000]
Controls	✓	✓	✓	✓	✓
Experiment: Hero-Hero	✓				
Experiment: Hero-Hero-Villain		✓	✓		
Experiment: Villain-Villain-Hero				✓	✓
Mean Outcome Control Group	1.35	1.25	0.67	1.01	0.81
Observations	789	798	798	699	699

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' memory of characters divided by role. Column 1 includes only the Hero–Hero experiment and tests recall of Green Tech and US People (both heroes). Columns 2 and 3 include the Hero–Hero–Villain experiment, showing effects on recall of Green Tech and Regulations (heroes, column 2) and Fossil Industry (villain, column 3). Columns 4 and 5 cover the Villain–Villain–Hero experiment, showing effects on recall of Green Tech (hero, column 4) and Fossil Industry/Corporations (villains, column 5). We use robust standard errors. Figure 12c in the paper is the reference plot.

F.3 Output with Randomization Inference

In this section we present robustness tests for the experimental results. In particular we reproduce all the results of the paper, computing the p-values via randomization inference, using the *ritest* Stata package.

Table F.9: ***Manipulation Check - Effectiveness of the Political Narrative Treatment - Randomized Inference***

Dependent Variable	**Mention of Character-Role**							
	Coeff./SE/ *Randomized (n=1000)* p-value							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.293	0.281	0.254	0.271	0.358	0.322	0.280	0.316
	(0.032)	(0.032)	(0.031)	(0.018)	(0.034)	(0.035)	(0.032)	(0.019)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Outcome | Character					✓	✓	✓	✓
Experiment: Hero-Hero	✓			✓	✓			✓
Experiment: Hero-Hero-Villain		✓		✓		✓		✓
Experiment: Villain-Villain-Hero			✓	✓			✓	✓
Experiment FE				✓				✓
Mean Outcome Control Group	0.33	0.34	0.44	0.37	0.45	0.52	0.64	0.54
Observations	987	976	968	2,931	742	686	695	2,123

Notes: The table reports OLS estimates from manipulation checks of the narrative treatment. The dependent variable is a binary indicator equal to one if the participant recalled the role assigned to a character in the treatment tweet (e.g., Green Tech as hero in the Hero–Hero experiment, Fossil Industry as villain in the Hero–Hero–Villain experiment). Columns 1–4 report effects on the unconditional likelihood of recalling the character-role, while Columns 5–8 restrict the sample to participants who recalled at least one character from the tweet. Only the treatment group was exposed to characters explicitly framed in roles; control participants saw the same characters presented neutrally. Non-zero recall in the control group therefore reflects participants' prior beliefs or participants' interpretation of the control tweet, while the treatment effect captures the additional role attribution induced by framing. All regressions control for income, education, political preference, age, and sex. Standard errors are clustered at the individual level. We compute the p-values using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. Table G.16 in the paper is the reference table.

Table F.10: ***Experimental Results - The Impact of Political Narratives on Beliefs - Randomization Inference***

Dependent Variable	**Expectation for the Future:**					
	Forecast	**Confidence**	**Forecast**	**Confidence**	**Forecast**	**Confidence**
	Coeff./SE/*Randomized (n=1000)* p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	2.487	3.426	0.831	0.331	-5.022	-0.358
	(1.368)	(1.680)	(1.425)	(1.740)	(1.387)	(1.803)
	[0.077]	[0.051]	[0.546]	[0.860]	[0.000]	[0.835]
Controls	✓	✓	✓	✓	✓	✓
Experiment: Hero-Hero	✓	✓				
Experiment: Hero-Hero-Villain			✓	✓		
Experiment: Villain-Villain-Hero					✓	✓
Mean Outcome Control Group	39.02	46.02	43.67	48.54	42.32	48.15
Observations	987	987	976	976	968	968

Notes: The table displays the coefficients of OLS regression models providing insights on two outcomes: the participants' forecast, as answer to the question *'What percentage of US energy do you predict will come from renewable sources and green technology by the year 2035? Indicate a number between 0 and 100.'* (in Columns 1, 3, 5), their confidence in the forecast, as answer to *'Your response on the previous screen suggests that by 2035, [x]% of US energy will come from renewable sources and green technology. How certain are you that the actual share of renewable energy in 2035 will be between [x-5] and [x+5]%?'* (in Columns 2, 4, 6). Columns 1 and 2 show results for the Hero-Hero experiment, columns 3 and 4 for the Hero-Hero-Villain experiment, and 5 and 6 for the Villain-Villain-Hero experiment. All models include income, education, political preference, age, and sex as controls. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. Figure 10a in the paper is the reference plot.

Table F.11: ***Experimental Results - The Impact of Political Narratives on Stated Preferences - Randomized Inference***

Dependent Variable	**Stated Preferences: Policy Support**		
	Coeff./SE/ *Randomized (n=1000)* p-value		
	(1)	(2)	(3)
Treatment	0.061	0.054	0.094
	(0.090)	(0.086)	(0.117)
	[0.501]	[0.558]	[0.474]
Controls	✓	✓	✓
Experiment: Hero-Hero	✓		
Experiment: Hero-Hero-Villain		✓	
Experiment: Villain-Villain-Hero			✓
Mean Outcome Control Group	5.70	5.72	4.19
Observations	987	976	968

Notes: The table displays the coefficients of OLS regression models providing insights on the support or opposition for a policy or law that is in line with the content of the narrative. In column 1, we show results for the Hero-Hero experiment, where participants were asked whether they would support a policy that reduces the cost of residential renewable systems. In Column 2 we show results for the Hero-Hero-Villain experiment, where participants were asked whether they would support increasing transparency and accountability for energy companies. In Column 3, we show results for the Villain-Villain-Hero experiment, where people were asked whether they would support raising taxes on fossil fuels. All models include income, education, political preference, age, and sex as controls. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. Figure 10b in the paper is the reference plot.

Table F.12: ***Experimental Results - Impact of Political Narratives on Revealed Preferences - Randomized Inference***

Dependent Variable	**Revealed Preference: Incentivized Donation**			
	Coeff./SE/*Randomized (n=1000)* p-value			
	(1)	(2)	(3)	(4)
Treatment	0.734	0.530	0.584	0.562
	(0.497)	(0.469)	(0.480)	(0.272)
	[0.150]	[0.263]	[0.216]	[0.045]
Controls	✓	✓	✓	✓
Experiment: Hero-Hero	✓			✓
Experiment: Hero-Hero-Villain		✓		✓
Experiment: Villain-Villain-Hero			✓	✓
Experiment FE				✓
Mean Outcome Control Group	6.52	6.40	6.73	6.55
Observations	987	976	968	2,931

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' revealed preferences. We measure revealed preferences with the decision to donate to an association promoting sustainable development and local/national projects to support Green tech diffusion. The decision is incentivized with a lottery: participants could be selected to win 25$ and had to allocate the amount between themselves and the association. No restrictions on the allocation were given. Columns 1, 2, and 3 show results for the single experiments, Column 4 shows results for the pooling together all experiments, and it includes experiment fixed effects. All models include income, education, political preference, age, and sex as controls. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. Figure 11 in the paper is the reference plot.

Table F.13: ***Experimental Results - The Impact of Political Narratives on Memory (Pooled Sample) - Randomized Inference***

Dependent Variable	**Information Retention:**			
	Facts		**Character**	
	Coeff./SE/*Randomized (n=1000)* p-value			
	(1)	(2)	(3)	(4)
Treatment	0.001	-0.002	0.061	0.049
	(0.012)	(0.013)	(0.016)	(0.021)
	[0.969]	[0.906]	[0.000]	[0.014]
Controls	✓	✓	✓	✓
Experiment FEs	✓	✓	✓	✓
Day of the Experiment	✓		✓	
Day After the Experiment		✓		✓
Mean Outcome Control Group	0.13	0.10	0.69	0.45
Observations	2,931	2,280	2,931	2,269

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' memory. Columns 1 and 2 show the effect on information retention, respectively in the day of the main experiment and a day later. Participants were asked to remember the factual information reported in the tweet. Column 3 and 4 show the effect on recalling the characters present in the text. All regressions include experimental FEs and the following controls: income, income, education, political preference, age, and sex. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. Paper Figure 12a and Figure 12b are the reference plots.

Table F.14: ***Experimental Results - The Impact of Political Narratives on Memory of Roles - Randomized Inference***

Dependent Variable	**Information Retention:**				
	Hero	**Hero**	**Villain**	**Hero**	**Villain**
	Coeff./SE/p-value				
	(1)	(2)	(3)	(4)	(5)
Treatment	0.169	-0.106	0.133	-0.139	0.636
	(0.075)	(0.075)	(0.059)	(0.064)	(0.077)
	[0.025]	[0.156]	[0.025]	[0.032]	[0.000]
Controls	✓	✓	✓	✓	✓
Experiment: Hero-Hero	✓				
Experiment: Hero-Hero-Villain		✓	✓		
Experiment: Villain-Villain-Hero				✓	✓
Mean Outcome Control Group	1.35	1.25	0.67	1.01	0.81
Observations	784	792	792	693	693

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' memory of characters divided by role. Column 1 includes only the Hero–Hero experiment and tests recall of Green Tech and US People (both heroes). Columns 2 and 3 include the Hero–Hero–Villain experiment, showing effects on recall of Green Tech and Regulations (heroes, column 2) and Fossil Industry (villain, column 3). Columns 4 and 5 cover the Villain–Villain–Hero experiment, showing effects on recall of Green Tech (hero, column 4) and Fossil Industry/Corporations (villains, column 5). All models include income, education, political preference, age, and sex as controls. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. Figure 12c in the paper is the reference plot.

G Political Narratives Correlation with Survey Data: Cooperative Election Study

Table G.15: ***Correlation Between Narratives on Green Tech and Public Preferences for Renewable Energies (CES Survey Data)***

Dependent Variable	**Preference Pro Renewable Energies**					
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Green Tech-Hero	0.019	0.019	0.018			
	(0.018)	(0.018)	(0.018)			
	[0.283]	[0.283]	[0.322]			
Share of Green Tech-Villain				-0.489	-0.489	-0.489
				(0.106)	(0.106)	(0.153)
				[0.000]	[0.000]	[0.001]
State FEs	✓	✓	✓	✓	✓	✓
Time Trend		✓	✓		✓	✓
Personal Characteristics			✓			✓
Observations	308495	308495	308495	308495	308495	308495

Notes: The table displays coefficients from Linear Probability Regression Models estimating the correlational impact of Political Narratives about Green Tech on public preferences for renewable energy use. The dependent variable is based on responses to the following question from the Cooperative Election Study survey: *'Do you support or oppose each of the following proposals? Require that each state use a minimum amount of renewable fuels (wind, solar, and hydroelectric) in the generation of electricity even if electricity prices increase a little'*. The question covers the period 2014-2021 and remains at the individual level. The independent variable is the state-year share of narratives about Green Tech framed either as Hero or Villain. Columns (1) and (4) include a yearly time trend; columns (2) and (5) additionally control for personal characteristics (gender, race, age, and education); and columns (3) and (6) further include state fixed effects. All regressions use sampling weights to ensure representativeness of the US population. Robust standard errors are employed.

Table G.16: ***Correlation Between Narratives on Regulations and Public Preferences on the Environmental Protection Agency (CES Survey Data)***

Dependent Variable	**Index Pro Environmental Protection Agency**					
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Regulations-Hero	-0.066	-0.066	-0.064			
	(0.056)	(0.056)	(0.055)			
	[0.243]	[0.243]	[0.242]			
Share of Regulations-Villain				-0.088	-0.088	-0.244
				(0.063)	(0.063)	(0.089)
				[0.159]	[0.159]	[0.006]
State FEs	✓	✓	✓	✓	✓	✓
Time Trend		✓	✓		✓	✓
Personal Characteristics			✓			✓
Observations	307690	307690	307690	307690	307690	307690

Notes: The table displays coefficients from Linear Probability Regression Models estimating the correlational impact of Political Narratives about Regulations on an index of public preferences regarding the Environmental Protection Agency. The dependent variable is based on responses to the following questions from the Cooperative Election Study survey: *'Do you support or oppose each of the following proposals? - Give the Environmental Protection Agency power to regulate Carbon Dioxide emissions'* and *'Do you support or oppose each of the following proposals? - Strengthen the Environmental Protection Agency enforcement of the Clean Air Act and Clean Water Act even if it costs US jobs'*. Responses to these questions are averaged into a single index covering the period 2014-2021 and remain at the individual level. The independent variable is the state-year share of narratives about Regulations framed either as Hero or Villain. Columns (1) and (4) include a yearly time trend; columns (2) and (5) additionally control for personal characteristics (gender, race, age, and education); and columns (3) and (6) further include state fixed effects. All regressions use sampling weights to ensure representativeness of the US population. Robust standard errors are employed.

H Political Narratives in Other Media

The main analysis in this paper investigates the virality of narratives on social media, focusing on Twitter/X. While these results provide detailed insights into the determinants of virality in the context of social media engagement, it is important to consider whether similar patterns in the use and distribution of narratives extend beyond the online environment. This appendix addresses this question by exploring the presence of political narratives – as defined in this study – in traditional media sources, specifically newspapers and television.

This exercise serves two purposes. First, it provides external validation by examining whether the types of narratives that drive engagement on social media are also present in other influential media platforms. Second, it assesses the potential scope extension of our findings, recognizing that while virality is a unique feature of social media, the content and prevalence of narratives may reflect broader patterns in public discourse. We present a simple descriptive analysis of narrative distribution in newspapers and television during the same period covered by our main study. Although we cannot reproduce the virality results in these settings, identifying similar narrative patterns across media contributes to a more comprehensive understanding of the role narratives play in shaping political communication.

H.1 Newspapers

In preparing and classifying the newspaper articles, we follow as closely as possible the procedure to prepare the tweets. We source a random and representative set of newspaper articles. To that end, we use the three most widely circulated newspapers in the US; The New York Times, The Wall Street Journal, and USA Today. We download the articles from Factiva. For each newspaper, we download 3,000 articles for the period between 2010 and 2021. We download the articles in four-year intervals. To ensure a balanced distribution of popular articles over time for each newspaper, we source the 333 and 334 most popular articles from 2010-2013, then 2014-2017, and 2018-2021. We use exactly the same list of keywords (Oehl, Schaffer, and Bernauer 2017)(see Section A.1 for the full list of keywords). We minimally adapt the prompts to classify the tweets, changing the references from tweets to newspaper articles.

We find that the distribution of articles closely mirrors the overall distribution of articles from Twitter. The three most frequently recurring character-roles for human characters are US democrats heroes with 7.52%, corporations-villains with 6.33%, and US republicans with 6.14%. Similarly, the three most frequently recurring character-roles for instrument characters are US fossil industry-villain with 14.98%, green tech-hero with 9.92%, and regulations-hero with about 6%.

This resembles the ranking for instrument characters.

Table H.1: ***Share of Character-Roles in Relevant Newspaper Articles (United States, 2010-2021)***

Panel A: Human Characters

	Hero	Villain	Victim	Neutral	Total
Developing Economies	0.51	0.78	3.16	0.85	5.29
US Democrats	7.52	0.44	0.17	1.67	9.80
US Republicans	0.19	6.14	.	1.92	8.24
Corporations	1.92	6.33	0.17	10.64	19.05
US People	1.44	0.18	1.84	4.00	7.45

Panel B: Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emission Pricing	3.92	0.94	.	2.15	7.00
Regulations	5.96	1.23	.	3.71	10.90
Fossil Industry	0.08	14.98	0.06	2.18	17.29
Green Tech	9.92	0.36	.	2.80	13.08
Nuclear Tech	0.98	0.21	0.13	0.59	1.91

Notes: The table shows the frequencies of character-roles in the classified newspaper articles as a percentage of the total occurrences of each character. Shares are computed considering only the dataset of relevant snippets used in our analysis. We define an article as relevant if it features at least one character from our list. We include in the computation of shares only character roles that appear at least 100 times, thus excluding 'US Republicans-victim', 'Emission Pricing-victim', 'Regulations-victim', and 'Green Tech-victim', indicated by a dot in the tables. Panel (a) displays the shares for characters of the human type, while Panel (b) displays the same for characters of the instrument type. The column Neutral in both panels reports cases where the character is present in the article but is not depicted in one of the three specific roles. The occurrence of character-roles is not mutually exclusive, meaning multiple roles may appear in the same tweet.

H.2 Television

Table H.2: ***Share of Character-Roles in Relevant TV transcripts (Fox News and MSNBC, 2010-2021)***

Panel A: Human Characters

	Hero	Villain	Victim	Neutral	Total
Developing Economies	0.17	0.22	0.71	0.34	1.44
US Democrats	16.29	1.63	0.00	0.57	18.50
US Republicans	0.10	14.72	0.00	1.03	15.84
Corporations	1.51	7.97	0.00	1.74	11.23
US People	5.87	0.06	2.81	6.19	14.93

Panel B: Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emissions Pricing	2.01	2.23	0.00	3.09	7.33
Regulations	2.60	3.89	0.00	2.26	8.75
Fossil Industry	0.18	10.85	0.00	1.30	12.33
Green Tech	6.35	0.15	0.00	2.29	8.79
Nuclear Tech	0.20	0.07	0.00	3.42	3.68

Notes: The table shows the frequencies of character-roles in the classified tv transcripts as a percentage of the total occurrences of each character. Shares are computed considering only the dataset of relevant tv transcripts used in our analysis. We define a snippet as relevant if it features at least one character from our list. We include in the computation of shares only character roles that appear at least 100 times, thus excluding ' US Republicans-victim', 'Emission Pricing-victim', 'Regulations-victim', and 'Green Tech-victim', indicated by a dot in the tables. Panel (a) displays the shares for characters of the human type, while Panel (b) displays the same for characters of the instrument type. The column Neutral in both panels reports cases where the character is present in the tv transcript but is not depicted in one of the three specific roles. The occurrence of character-roles is not mutually exclusive, meaning multiple roles may appear in the same tweet.