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

Kai Gehring† and Matteo Grigoletto‡

June 18, 2026

Abstract

The effectiveness of 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 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. A pre-registered experiment with real-stakes sharing decisions across six non-climate topics replicates these effects causally and outside social media. 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 shifts beliefs and revealed preferences about a character, but not support for specific policies. Narratives also lead to better recall of the characters and their roles, while recall of objective facts is not improved. Taken together, the political narrative framework provides a measure that moves beyond emotions and linguistic features, helps explain virality, and links 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, Andrea Mentasti and Lorenz Gschwent for excellent research assistance; all remaining mistakes are our own. We are grateful for helpful suggestions from Toke Aidt, Elliott Ash, Bruno Caprettini, Milena Djourelova, Ruben Durante, Vera Eichenauer, Gaël Le Mens, Pierre-Guillaum Meon, Chris Roth, Paul Schaudt, Marta Serra-Garcia, Johannes Wohlfahrt, Joachim Voth, and many others, as well as from participants and discussants at the Barcelona Summer Forum Political Economy 2026 and Summer Forum Behavioral Foundations of Information, Attention, and Belief Formation 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, Lugano University, University of Otago, University of Queensland, University of New South Wales 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). We provide a package that operationalizes our pipeline, enabling users to extract political narratives across topics. You can access the package at the link: Political Narratives Package.

†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 so-called "drama triangle" (Karpman 1968) – 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, Haaland, et al. (2025), Eliaz and Spiegler (2020), and Kendall and Charles (2022) started analyzing narratives systematically as *causal narratives* – directed graphs of cause–effect links between nodes – and Barron and Fries (2024) evaluate their persuasive power. 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* – organized around characters and their drama-triangle roles – 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 political 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 pre-defined 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, with villain combinations consistently amplifying the effect. 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.

The following passage by climate activist Greta Thunberg illustrates how our approach defines

and captures the essence of a narrative:

> *"**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

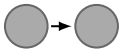
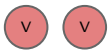
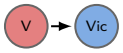
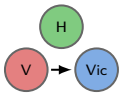
Corporations (*energy companies*), the people (*countless people*), and civil society (*younger people, activists*) are the characters in this passage. Most readers will agree on their positioning: corporations are cast as villain (red color throughout the paper), the poor as victim (blue color), and civil society as hero (green color). These role assignments are clear even though causal relationships between the characters are not explicitly stated and there is little use of emotions related to characters. Thinking of narratives as a dynamic graph, as introduced to economics by Eliaz and Spiegler (2020), characters constitute the nodes, roles are features of the nodes and causal claims the directed arrows between them.[1] The Thunberg passage shows that the characters and roles can be identified reliably and convey the essence of the narrative, even when causal arrows between the nodes are absent or ambiguous. Our definition and measurement of narratives therefore focuses on specifying the topic and characters, and coding for each character whether it appears as neutral or cast as hero, villain, or victim.

This approach allows clearly distinguishing whether a text contains a narrative or not, a necessary condition to analyze whether and which types of narratives go viral. Applying this distinction requires recognizing that on social media, in newspapers, and political speeches, narratives often appear as fragments that position one or two characters in a role without articulating the full causal chain. These fragments can be thought of as appealing to a meta-narrative that recipients are aware of (which often includes causal links between characters). Thunberg mentioning only CORPORATIONS-Villain in another tweet, for instance, will prompt recipients to think of the larger narrative — even though the fragment itself states neither the other characters nor any causal connection (see Figure C.6). Table 1 classifies texts along three dimensions — role assignment, sequences, and emotional language — to formalize what constitutes a narrative and what does not.

Row 1 provides an example of a neutral character (grey node), with neither a sequence nor emotional language — a statement that features a character but is clearly not a narrative. Row 2 features two characters linked by a causal arrow, yet neither is cast in a role — raising the question of whether a sequence alone always constitutes a narrative. Under a definition that treats causal structure as the sole defining feature, any text describing a causal relationship or temporal sequence would be labeled a narrative, including many scientific, merely logical or chronological statements. Rows 3 to 6 describe text types that we classify as narratives and illustrate how the same character as in row 1 and 2 is cast in a role (indicated by the colors of the respective nodes) through context, sequences, or emotions. This highlights how our approach provides a clear

[1] Nodes are also referred to as variables and arrows as "links". Roles as node features are specific to our framework.

Table 1: ***Definition of Political Narratives***

Type	Features	Visual	Example (Climate Policy)
1. No Narrative	No Roles; No Sequence *(No Emotion)*		"Recently, I became very curious regarding news about a 5% carbon tax."
2. No Narrative	No Roles; Sequence *(No Emotion)*		"A 5% carbon tax internalizes part of the social cost of carbon by affecting the price of fossil fuels."
3. Political Narrative	Roles; No Sequence *(No Emotion)*		"Clean Green Tech is the solution."
4. Political Narrative	Roles; No Sequence *(Emotion)*		"*Predatory* corporations thrive on *dangerous* fossil fuel extraction."
5. Political Narrative	Roles; Sequence *(No Emotion)*		"Price increases due to a 5% carbon tax raise the costs of living for Americans."
6. Political Narrative	Roles; Sequence *(Emotion)*		"Price increases due to a 5% carbon tax raise the cost of living of *vulnerable* Americans-clean Green Tech is the solution."

Notes: The table classifies narrative fragments based on the presence of character roles, causal sequences, and emotional triggers. Highlighting matches visual node roles: Villain (Red), Hero (Green), Victim (Blue), and Grey (Neutral). Circles in visual always represent characters. Emotion-charged words are indicated in italics.

counterfactual for empirical work and allows capturing narratives of differing complexities while being compatible with an understanding of narratives as causal sequences.

To measure political narratives with LLMs for any topic and user-defined set of characters, we provide a general pipeline, implemented in Python.[2] After filtering texts for the selected topic, the pipeline prompts an LLM to (i) flag the presence of pre-defined characters and (ii) assign drama triangle roles where present. The ubiquity of the archetypal roles in human texts ensures that both human coders with appropriate instructions and LLMs with simple one-shot prompting can identify role assignments reliably. Our pipeline returns a compact panel of variables – one row per text unit and columns indicating character presence and role assignment – that is straightforward to use for further analysis.

To analyze what makes 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

[2]The pipeline is available at github.com/KaliGi/Political-Narratives-Package.

influence. Because the underlying issue is characterized by deferred pay-offs and distributional conflict, 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 narratives evolve alongside the shifting public support patterns reported in cross-country surveys (Andre, Boneva, 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 use further metadata to select tweets originating from the US, ending with roughly 1.15 million tweets. We then apply our pipeline, using GPT-4o-mini from OpenAI, to validate topic relevance, detect characters, and assign roles for each in-topic tweet. In a 500-tweet validation exercise, LLM-to-human agreement on character presence and roles is comparable to inter-human agreement, while the LLM substantially outperforms a traditional NLP baseline combining dictionary methods with dependency parsing.

Descriptive results provide an initial overview of the climate-policy discourse. Green Tech emerges as the dominant hero, Fossil Industry as a key villain, and US People as the most frequent victim. US People also appears frequently as hero, illustrating that the same character takes different roles across tweets, and character-role frequencies shift markedly over the 2010-2021 period, only partly reflecting real-world policy developments. Among common combinations, Green Tech-Hero paired with Fossil Industry-Villain and Fossil Industry-Villain paired with Corporations-Villain are among the most frequent. The majority of tweets features one or two character-roles, and more than three character-roles are the exception. For the systematic analysis of what drives virality, we focus on the overarching role categories hero, villain, and victim, to compare narratives to tweets featuring only neutral characters and to examine which roles and role combinations drive the effect

We exploit the fact that the retweet metric on Twitter constitutes a natural proxy for virality.[3] 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

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

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 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 virality, 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 virality.

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. 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. Fourth, our analysis of virality is correlational: despite extensive controls and fixed effects, selection and confounding cannot be entirely ruled out. Fifth, 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. Finally, 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 directly address the most important of these limitations, we conduct a pre-registered virality experiment with 600 participants from a representative US population. Each participant makes six sharing decisions in random order; for each text, they randomly see either the narrative or the neutral version and decide whether to share it with another Prolific participant. Sharing is a real-stakes decision: shared texts are forwarded to another Prolific user. The experiment uses topics outside climate change policy, providing a direct test of external validity. We find that narratives causally increase sharing: hero and villain narratives significantly boost the probability of sharing, while

victim narratives do not, mirroring the observational results. While more subtle details or effect sizes will be context-dependent, we conclude that the main patterns identified with our narrative framework seem to hold for more general settings and are not driven by algorithmic amplification or automated accounts, concerns inherent to observational social-media research.

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 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. 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 narrative exposure does not significantly shift concrete policy preferences, but we cannot rule out that repeated exposure to highly viral narratives on social media would eventually shift policy preferences (see review in Montoya et al. 2017). Finally, 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.

We thus make five core contributions. First, we develop the political narrative framework – a formal, topic-agnostic approach that defines narratives by characters and their roles rather than by causal sequences – and operationalize it through a reusable LLM pipeline. Second, we provide the first systematic evidence on narrative virality in economics, showing which character-role combinations drive spread and confirming these patterns causally in a pre-registered experiment. Third, we show that single exposure to political narratives shifts beliefs and a real-stakes donation choice. Fourth, we identify a selective memory channel through which drama triangle roles enhance character recall, with villain roles producing the strongest effect. Finally, we contribute to media economics and the economics of social media by shifting the focus from media exposure to the structure of the content that spreads. We discuss these contributions and the related literature in more detail in Section 2.

2 Contributions and Related literature

We develop the political narrative framework, a formal approach that defines a narrative by its topic, its characters, and at least one character cast as hero, villain, or victim – the most parsimonious set of roles preserving the distinction between emotionally comparatively negative active agents (villains) and passive agents (victims).[4] The economic analysis of narratives has so far focused on causal sequences — modelling narratives as directed graphs (Eliaz and Spiegler 2020), measuring

[4]Character-role taxonomies have a long tradition in political science (Jones and McBeth 2010), sociology (Polletta et al. 2011), and communication studies (Anker 2005), but have remained almost entirely qualitative. Gehring, Adema, and Poutvaara (2022) show that simply measuring positive vs. negative sentiment without accounting for victim roles would create a biased evaluation of immigrant narratives in German newspapers. Algan et al. (2025) demonstrate the importance of emotions, but do not link them directly to characters or distinguish between villains and victims.

them via manual coding (Andre, Haaland, et al. 2025), or formalizing causal narrative theory (Kendall and Charles 2022). In real texts, people often do not specify causal links explicitly. Our approach captures more complex narratives as well as fragments of grander meta-narratives, and causal links between nodes can be estimated on top if specified in a text (see Subsection C.3). We operationalize it through a reusable Python/LLM pipeline that produces structured character-role panels, moving dimension reduction upfront rather than deferring it as in tools like RELATIO (Ash, Gauthier, and Widmer 2024).[5] The framework is topic-agnostic and portable: whenever a topic and a set of salient characters can be specified, researchers can apply the same scheme.

Second, we apply the framework to show that narratives are substantially more viral than comparable neutral texts and identify which roles, character types, and combinations drive this effect — providing, to the best of our knowledge, the first systematic evidence on narrative virality in economics. A pre-registered experiment with real-stakes sharing decisions confirms these patterns causally and extends them beyond climate policy. This augments a growing observational literature showing that specific narratives causally affect high-stakes behavior, from revisionist villain narratives that reshaped political preferences after the US Civil War (Esposito et al. 2023) to COVID narratives that changed health behavior conditional on who is blamed as the villain (Bursztyn et al. 2023). These studies each examine one particular narrative; our framework enables a comparative analysis across character-roles and their combinations. Virality itself has received little empirical attention in economics; the few studies outside the discipline examine emotions and arousal (Berger and Milkman 2012) or truth value (Vosoughi, Roy, and Aral 2018) rather than narrative structure.[6]

Third, moving beyond virality, our pre-registered survey experiments show that role-based narratives persuade: single exposure shifts beliefs and a real-stakes donation choice about the named character, even without an explicit causal chain. Andre, Haaland, et al. (2025) show that explicit causal sequences shift beliefs, and Barron and Fries (2024) identify sense-making and fit to facts as key persuasion mechanisms. These studies test highly structured stimuli, but real-world narratives often take the simpler form of role assignments rather than articulated causal reasoning. Our evidence brings role-based persuasion into a controlled experimental setting, closer to how narratives actually appear in political communication than the structured stimuli tested in prior work (e.g. Graeber, C. Roth, and Zimmermann 2024). In their review of information provision experiments, Haaland, C. Roth, and Wohlfart (2023) note that narratives represent a promising but underexplored intervention format; our experiments provide direct evidence that this format shifts both beliefs and behavior, complementing the broader persuasion literature (reviewed in DellaVigna and Gentzkow 2010).

Fourth, we identify a selective memory channel that might partly explain the persuasion effects:

[5] Similar to other recent papers employing LLMs or deep-learning models (e.g. Ash, Durante, et al. 2021; Lagakos, Michalopoulos, and Voth 2025; Voth and Yanagizawa-Drott 2025) to encode stories and images, we demonstrate how LLMs allow moving beyond simple metrics like emotions to study structural features of narratives systematically.

[6] Our results are consistent with the political economy of climate change literature evidence that misperceived social norms affect climate policy support (Andre, Boneva, et al. 2024), that media coverage of natural disasters shapes climate beliefs (Djourelova, Durante, Motte, et al. 2024), and that ideology outweighs knowledge in shaping climate preferences (Dechezleprêtre et al. 2025).

characters and their roles are better recalled and villain character memory crowds out hero memory, while recall of numerical facts embedded in the same narratives is unaffected. Consistent with the emerging literature on attention economics (Serra-Garcia 2025; Loewenstein and Wojtowicz 2025), drama triangle roles could secure attention by activating cognitive heuristics that simplify complex political information — with villain characters, who pose a perceived threat, receiving disproportionate processing. Combined with the virality finding, this creates a reinforcing mechanism: narratives spread further, persuade more and are remembered better. These results open space for follow-up research on how humans use role-based heuristics to navigate information-rich environments, and how senders may strategically exploit such patterns to shape what audiences remember.

Fifth, these findings contribute to media economics (see overview in Zhuravskaya, Petrova, and Enikolopov 2020) and the economics of social media (see overview in Aridor et al. 2024) by shifting the focus from media *exposure* to the *structure* of the content that spreads. A large body of work has established that exposure to media content — whether newspapers (Djourelova, Durante, and Martin 2025), television (e.g., Ash, Galletta, et al. 2024; Ash and Poyker 2024; Enikolopov, Petrova, and Zhuravskaya 2011; Durante, Pinotti, and Tesei 2019), radio (e.g., Yanagizawa-Drott 2014), or folklore and movies (Michalopoulos and Xue 2021; Michalopoulos and Rauh 2024) — shapes beliefs, preferences, and behavior. Work on social media shows that online platforms can influence offline outcomes including protest participation (Enikolopov, Makarin, and Petrova 2020) and hate crimes (Müller and Schwarz 2021; Müller and Schwarz 2023). Yet why certain messages reach large audiences in the first place has received far less attention. We provide a specific answer: character-role assignments and their combinations are a structural feature of human language that predicts which narratives go viral, moving beyond proxying content by topic keywords or sentiment (e.g. Braghieri et al. 2024).

3 Definition and Measurement of Narratives

3.1 Definition

Our definition of narratives complements the dominant "narratives -- as–causal–sequences" approach in economics, while drawing on approaches that regard narrative more broadly as communicative devices to gain attention, encode roles and identities, and thereby shape 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, nation 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 $r(k) \in \{\text{hero}, \text{villain}, \text{victim}\}$. This definition accommodates fragments and non-sequential formulations (e.g., "Corporations are villains") while remaining compatible with causal or temporal representations (Subsection C.3).

The drama-triangle roles are the most parsimonious set that still captures essential differences across roles. Reducing the number of roles further would yield a simple positive-negative polarity, missing the key distinction between active agents as villains and a passive agent as victims. A sentence like "It is horrible how Corporations exploit Workers" is overall negative regarding both characters, yet the narrative logic crucially depends on who is the villain and who the victim.[7] For these reasons the drama-triangle roles provide a recurrent role scheme that is widely reflected in human story telling from biblical parables and classical epics to contemporary news and social media.

The formal notation also helps to clarify our 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 the text makes causal relations explicit, one can estimate causal relationships $E \subseteq K' \times K'$ on the same K, layering a causal graph on top of the character-role coding as an optional downstream step that presupposes the consistent node definition.

Finally, the same logic is portable beyond climate-policy text. For example considering monetary policy as a topic, 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.

[7] 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, through 7 in Propp's morphology, to 12-16 archetypes according to Carl Jung. Many, though not all, of these more distinct identities and roles can be meaningfully collapsed into the three drama triangle roles (see evidence in Bergstrand and Jasper 2018). Most alternative role systems explicitly include one or more victim roles (e.g. Propp and Jung).

3.2 Pipeline

Let $\mathcal{D}$ be the set of documents, K the set of user-defined characters for the topic (as defined in Section 3), and $\mathcal{R} = \{\text{hero}, \text{villain}, \text{victim}\}$. The pipeline produces (i) a document-by-character presence matrix $\mathbf{M} = (m_{ik})_{i\in\mathcal{D},k\in K}$ with $m_{ik} \in \{0,1\}$ and (ii) role indicators $\mathbf{R} = (r_{ikr})_{i\in\mathcal{D},k\in K,r\in\mathcal{R}}$ 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. Thus $r_{ikr} = 1$ if and only if $r(k) = r$ in document i, linking the indicators to the role-assignment function defined in Section 3.

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. Figure 1 illustrates the five implementation steps.

1. Select and define the topic. The topic T fixes the boundary of relevance and anchors the character set K.

2. Identify the source and extract data. For a topic T, dictionaries and Boolean keyword queries are an efficient way to narrow a large text corpus to a topic-focused dataset. Further useful pre-processing steps can include language filtering, de-duplication, and optional geo-filtering.

3. Identify relevant characters. Character selection and definition map the topic into a manageable set of human and instrument characters. 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`/`relevant`); (ii) conditional on `relevant`, character presence and role assignment over $K \times \mathcal{R}$.

5. Obtain predictions and assemble outputs. We process documents via API calls, parse JSON responses, and assemble a tidy panel with (i) relevance classification, (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. Quality depends on the precise definition of topic, on initial text selection, and on the precision of the character descriptions, and it might also be context and culture dependent. Human validation is a costly, but necessary step to evaluate quality.

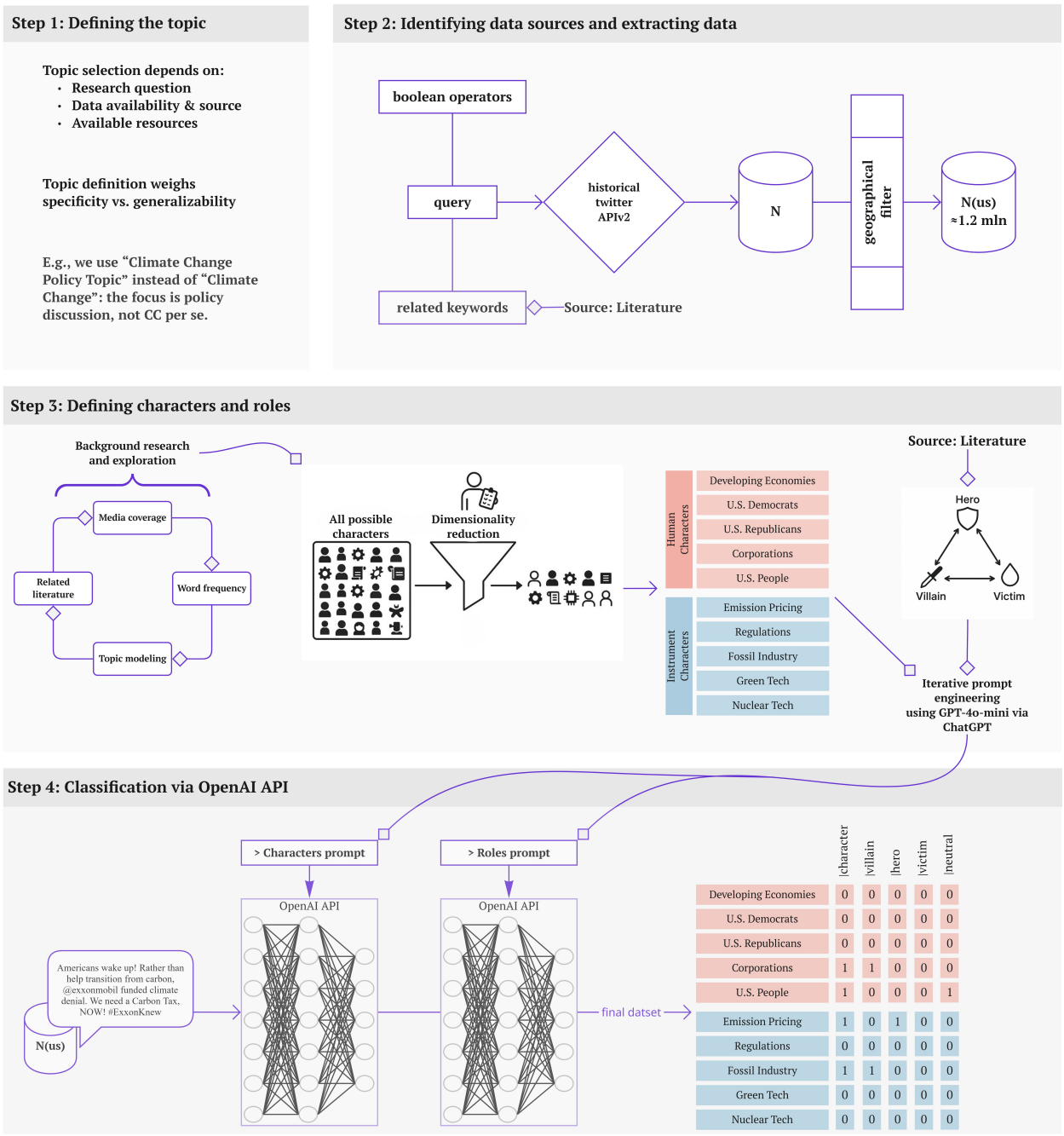
4 Data: Twitter Sample over 2010– 2021

This section applies the five-step pipeline; 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

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 text. The implementation parallelizes request preparation and uses API batch processing to scale to millions of tokens (see Subsection A.3).

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. The API's retweet count excludes quote-tweets, so our virality measure captures real endorsement. We further 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 1,151,521 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 3), 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}.[8] 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.

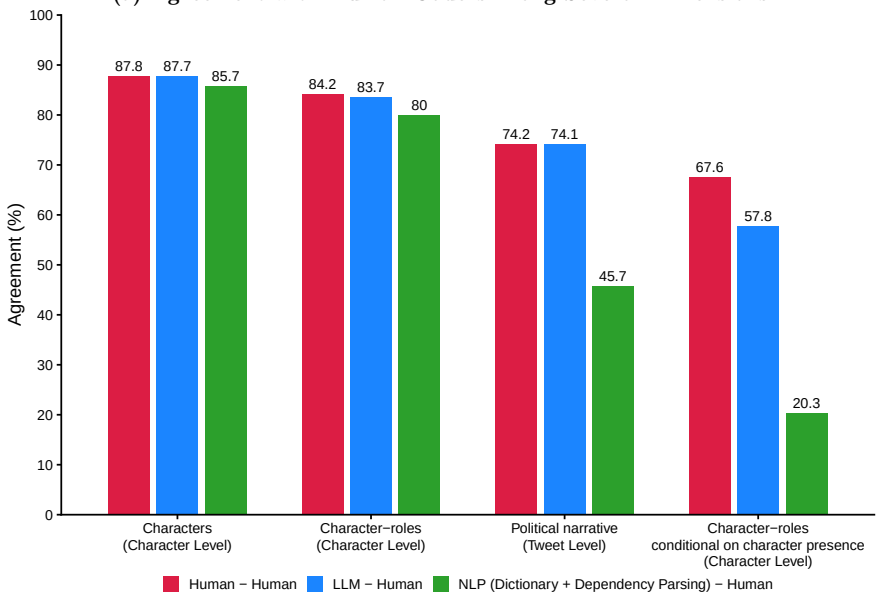
Quality control and validation. Unlike a mathematical question or fact-checking, narrative character detection and role assignment have no objective ground truth. Two human coders may agree that a character is present, but disagree on whether the character is cast as a hero, villain, or victim. For that reason, we randomly sample 500 relevant tweets and have each coded by two human coders (28 in total) recruited via Amazon Mechanical Turk, using the same role taxonomy and instructions as the LLM pipeline. Human intercoder reliability then serves as the natural benchmark: the LLM can be at most as good as humans are among themselves. In addition, we construct a dictionary-plus-parsing NLP baseline to compare against traditional text-as-data methods in economics (see Gentzkow, Kelly, and Taddy 2019). It uses named-entity recognition and LLMs to build keyword dictionaries, then applies syntactic dependency parsing to assign roles.[9] Details in Appendix B.

[8] Estimating directed causal relationships $E \subseteq K' \times K'$ between characters is a separate, optional step on the same node set, not part of our main analysis because only a subset of tweets contain explicit causal chains.

[9] We deliberately keep role assignment rule-based in the NLP baseline so the contrast isolates the value added by LLM-based inference, which traditional methods cannot replicate when role assignment depends on context, framing, and rhetorical intent.In Gehring and Grigoletto (2023) we separately tested a supervised pipeline combining a fine-tuned RoBERTa classifier with XGBoost, trained on 10,230 MTurk-annotated tweets. Per-character-role F1 scores were slightly lower and such an approach requires much more effort and higher costs for training.

Figure 2: ***Performance of LLMs Classification vs Human Coders and Alternative NLP Methods***

(a) ***Agreement with Human Coders Along Several Dimensions***



(b) ***Performance Across Tasks and Samples***

Panel	Task	Comparison with Human Coders	Accuracy	F1	Precision	Recall
A	Character presence	LLM	92.9	74.9	68.1	83.2
		Dictionary + Dependency Parsing	90.6	62.3	63.6	61.1
B	Character roles	LLM	69.8	59.9	75.8	53.6
		Dictionary + Dependency Parsing	19.9	20.2	38.6	20.5
C	Political narrative	LLM	82.5	89.9	90.6	89.2
		Dictionary + Dependency Parsing	44.2	53.3	98.3	36.5

Notes: Figure 2 reports agreement rates between pairs of human coders, between the LLM-based classifier and human coders, and between the dictionary + dependency parsing pipeline and human coders. We evaluate four dimensions: *Characters* (presence, including positive and negative cases; $N = 5{,}000$), *Character-roles* (exact role match, including absence; $N = 5{,}000$), *Political narrative* (agreement on whether at least one non-neutral role is present in a tweet; $N = 500$), and *Character-roles conditional on character presence* (role agreement conditional on human agreement that a character is present; $N = 558$). Figure 2b uses human coding as the ground truth and restricts evaluation to cases where the two human coders agree. Panel A evaluates character presence (absence vs. any role) conditional on human agreement on presence/absence ($N = 4{,}392$). Panel B evaluates exact character roles conditional on humans agreeing on the role assignment on top of character agreement ($N = 377$). Panel C evaluates political narrative at the tweet level—defined as the presence of at least one Villain, Hero, or Victim role (Neutral excluded)—conditional on human agreement on narrative presence ($N = 371$). Accuracy is defined as $(TP + TN)/(TP + TN + FP + FN)$, precision as $TP/(TP + FP)$, recall as $TP/(TP + FN)$, where TP denotes true positives, TN true negatives, FP false positives, and FN false negatives. $F1 = 2 \cdot \text{precision} \cdot \text{recall}/(\text{precision} + \text{recall})$. Table B.1 reports performance metrics at the role level. Subsection B.2 provides more detailed information.

Figure 2a compares agreement rates across three pairings (human–human, LLM–human, and NLP–human) along four evaluation dimensions. For character presence, character-role assignment, and tweet-level narrative detection, LLM–human agreement is virtually identical to human–human agreement (87.7% vs. 87.8%, 83.7% vs. 84.2%, and 74.1% vs. 74.2%, respectively). The NLP baseline tracks human coders reasonably well on character presence (85.7%) but drops sharply to 46.3% for narrative detection.

The fourth dimension is most relevant for the LLM vs. NLP comparison, focusing on cases where both humans agree whether a certain character appears.[10] The large gap underscores that traditional NLP methods can detect a character rather well, but role assignment requires interpreting context, framing, and rhetorical intent, which is where LLMs add particular value. Adopting the same approach of treating human-agreed labels as approximate ground truth, Figure 2b complements agreement rates with standard classification metrics and confirms the same pattern: the LLM performs well and clearly outperforms simpler traditional methods.[11]

5 Descriptive Results: Features and Distribution of Narratives

5.1 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 3.2, let $m_{ik} \in \{0,1\}$ indicate the presence of character $k \in K$ 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 consists of those with characters, either depicted as neutral or cast in a drama triangle role. Table C.5 summarizes character-role shares among `relevant` tweets. Panel (a) covers human characters and Panel (b) instrument characters.

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 hero (about 4%) and as victim (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.

[10] We display the value for LLM-human agreement, but it represents a very conservative and arguably biased assessment. The human–human comparison evaluates role assignment alone (since both humans already agree on character presence), whereas the LLM must match on both character detection and role assignment. In cases where the LLM would have classified the character as absent, it assigns no role and automatically disagrees with both humans.

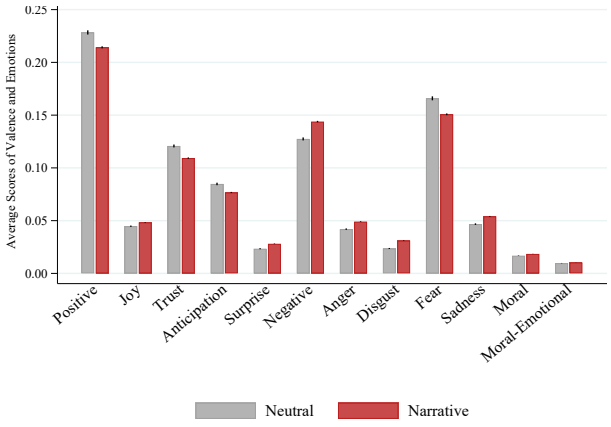
[11] The F1 scores for the LLM are within the range of papers in computer science on complex classification tasks (Salminen et al. 2020). Panel B shows that the LLM achieves an F1 of 59.9 with high precision (75.8%), indicating that our encoding tends to be conservative: it misses some assignments but rarely assigns incorrect ones. NLP is only superior in one comparison, and always inferior regarding overall accuracy or the F1-score.

5.2 Features of Political Narratives

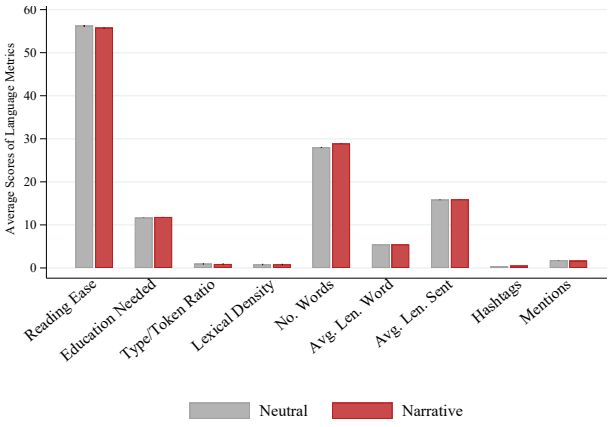
This section explores whether narratives differ systematically in emotional content/valence and in language metrics, two plausible determinants of virality and persuasion. Emotional content/valence is one plausible and common way of assigning roles, hence, we are interested in how roles differ with regard to the emotions they employ. Recently, economists have become more interested in emotions and the effect of specific emotions like anger (e.g. Algan et al. 2025). Writing style and quality are key features of a text, hence we want to explore whether there are systematic differences between narratives and neutral texts. Narratives may simply be easier to read or more emotionally charged than neutral tweets – features that, rather than the character roles, could drive virality and persuasion.

Figure 3: ***Valence, Emotions, and Language Metrics of Relevant Tweets (United States, 2010−2021)***

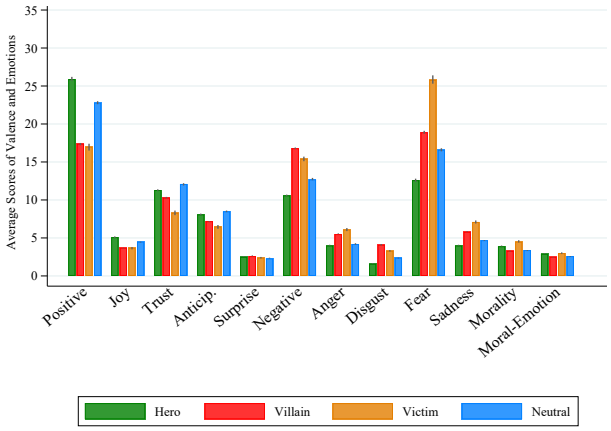
(a) ***Political Narrative vs. Neutral Characters: Emotions and Valence***



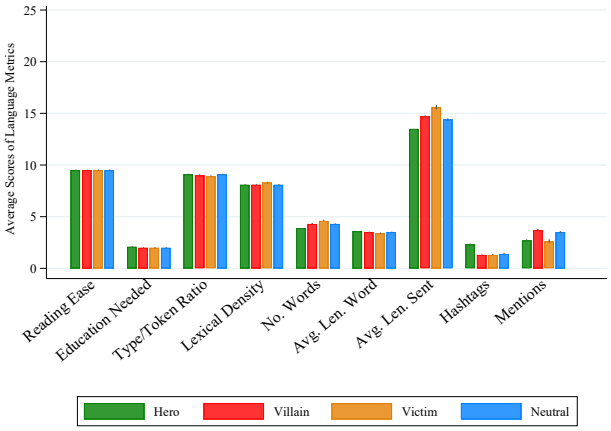
(b) ***Political Narrative vs. Neutral Characters: Language Metrics***



(c) ***Across Roles: Emotions and Valence***



(d) ***Across Roles: Language Metrics***



Notes: Panels 3a and 3b compare tweets with characters all in a neutral role to political narrative tweets (at least one $r_{ikr} = 1$). Panels 3c and 3d 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). Moral language measures use the Moral Foundations Dictionary (Graham, Haidt, and Nosek 2009).

Figure 3 summarizes the comparison. The top row contrasts political narratives tweets with tweets where all characters are featured neutrally; the bottom row differentiates between roles. Emotions and valence measures include those from the NRC Emotion Lexicon (Mohammad and Turney 2013) (positive, negative, joy, trust, anticipation, surprise, anger, disgust, fear, sadness), as well as the Moral Foundations Dictionary (Graham, Haidt, and Nosek 2009, moral) and an interaction of moral and emotional words that was shown to be related to retweet behavior (Brady et al. 2017). 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.

Three patterns emerge. First, political narratives tend to skew more negative in valence but there is no strong pattern across individual emotions overall. Second, role-level heterogeneity reveals a pattern that a simple positive–negative polarity would miss: hero narratives are the most positive; villain narratives are the most negative with anger and disgust particularly salient; victim narratives score negative in overall valence like villains but load overwhelmingly on fear. This emotion-level distinction between villains (anger/disgust) and victims (fear) is what our three-role taxonomy captures, which a binary sentiment measure cannot. Third, language metrics are broadly similar across political narratives and neutral tweets; a comparison between specific roles also only reveals a small difference with regard to average length.

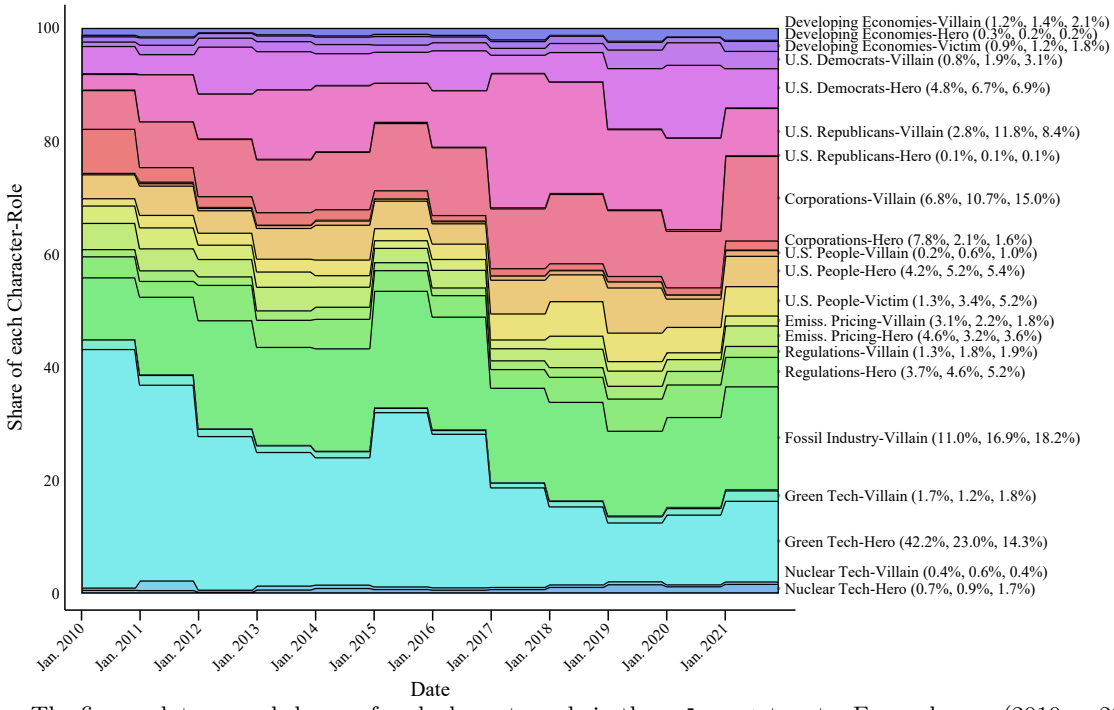
5.3 Political Narratives over Time

Compared to the pure frequency table by roles before, Figure 4 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 type's 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 a 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 Corporations appears to be linked to two big shifts in instrument characters. 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

Figure 4: ***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. Table C.7 is the reference table for this figure.

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 relative prominence 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 populism and Donald Trump, whose rhetoric frequently dramatizes political discourse through the use of the drama triangle roles. The reallocation towards human characters at the expense of instrument 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.

5.4 Character-role Combinations

The versatility of the political narrative framework allows us to move beyond marginal frequencies and investigate specific character-roles (CR) combinations as a heatmap in Figure 5. This offers a more granular view of the narrative structures in the corpus: where Figure 4 tracks how attention to individual character-roles evolves over time, the heatmap shows how character-roles combine within tweets. To keep the heatmap interpretable, we restrict to tweets with one or two character-roles. Diagonal cells show single-role political narratives; off-diagonal cells show character-role pairs. Shares are normalized by the total number of tweets in this subset, details in Subsection C.2. This approach allows us to both highlight the top 10 and top 30 most frequent combinations in different colors and to contrast single appearances with combinations.

While single occurrence as a narrative fragment is the most common way a character-role appears, a few specific CR combinations are among the most frequent narratives. 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.2). 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 among the top 30.[12] For each of these, retrieving the CR combination and studying some examples of such tweets gives a good indication about the essence of this type of narrative. Some CRs never occur together as a combination: for instance, the combination Fossil Industry-Hero and Emission Pricing-Hero. Taken together, this validates the importance of measuring fragments and shows how CR combinations can often preserve the essence of a narrative in a meaningful way.[13]

6 Results: Virality of Narratives on Social Media

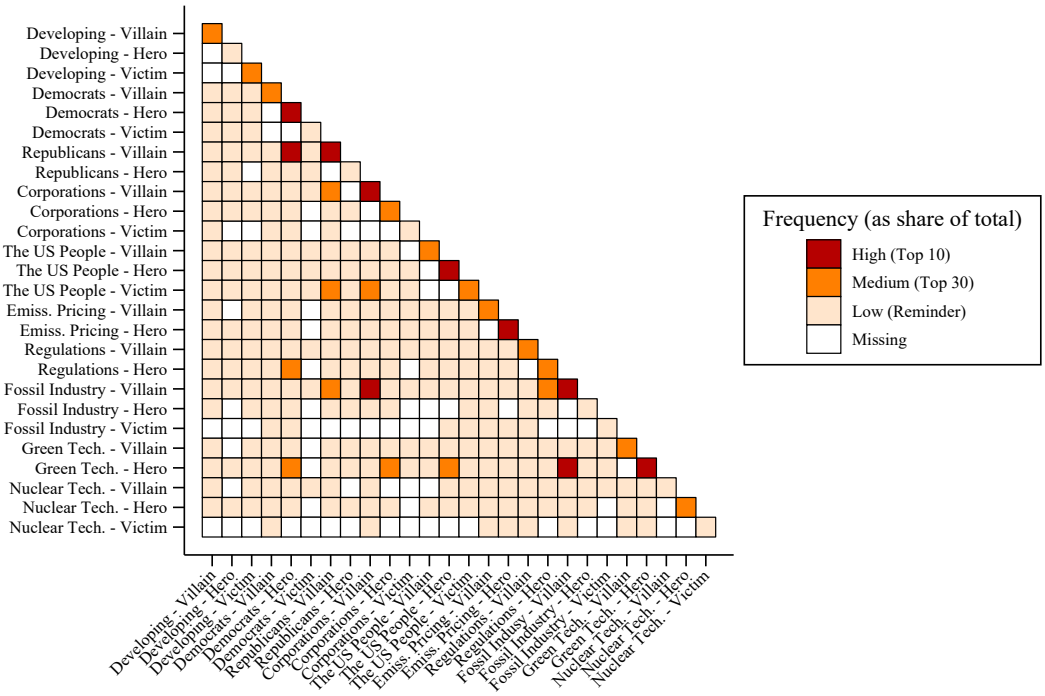
6.1 Virality of Character-Roles

Before turning to the main empirical results, we can descriptively examine variation in narrative virality across all one- and two-role combinations. Figure C.4 the character-role combinations most frequently composed by users are not always the ones that receive the most retweets, but there are some re-occurring patterns. Among the top 30 most viral combinations, Fossil Industry-Villain

[12] Political parties appear prominently as a partisan mirror (US Republicans-Villain paired with US Democrats-Hero, also among the top 10), but we defer a more detailed analysis of partisan narrative framing to future work (cf. Figure C.4).

[13] As an extension, we also an additional prompt that aims to understand in what share of cases we can measure causal connections between the characters (nodes) for all narratives with two (three) characters. Figure C.8 shows that in 34.5% (49.5%) we can measure at least one causal connection.

Figure 5: ***Absolute Frequency of Character-Roles Combinations-Tweets with One or Two Character-Roles***

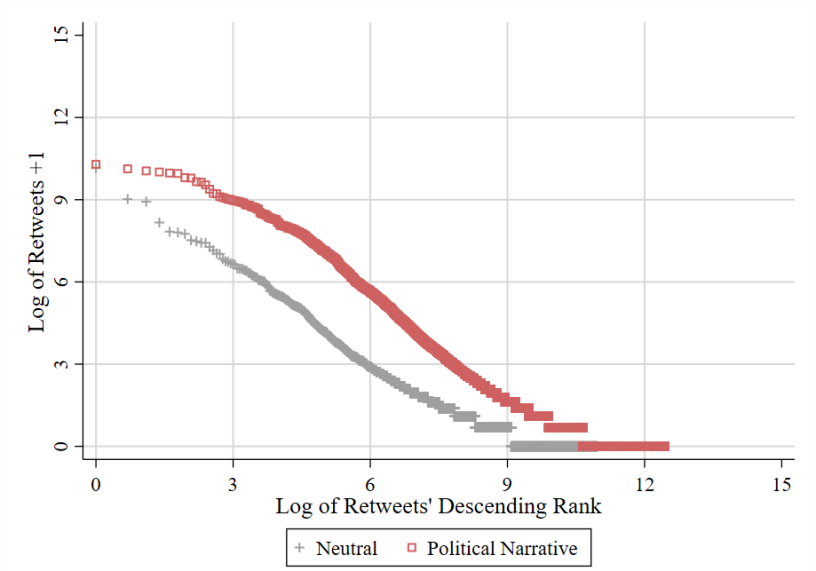


Notes: The figure shows the frequency of each character-role appearing alone or in combination with another character-role divided by all political narrative tweets with one or two character-roles. The diagonal of the matrix shows how often each character-role appears alone in a tweet. Tweets with three or more character-roles are excluded. Missing indicates either character-roles that appear less than 100 times over the entire 2010-2021 period or are by design, as every character can only appear in one role within one specific tweet. To 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.2 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.

appears in four different combinations and as a single fragment. Specific combinations include Fossil Industry-Villain with Corporations-Villain as a villain-villain combination featured in the top 5, as well as US People-Hero & Green Tech-Hero and the antagonistic combination Green Tech-Hero vs. Fossil Industry-Villain among the top 30. The US People character appears three times as hero and three as victim in the top 30 (see Figure C.4 for details).

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 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

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. The vertical position at a given rank reflects relative virality: being on a higher curve means tweets at that rank receive relatively more engagement.

against climate change by drafting and enacting important regulations. 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 lies above the neutral curve across the entire distribution, and the vertical gap widens toward the more viral end. Narrative tweets are 1.6× as likely as neutral ones to exceed the 100-retweet threshold; this rises to 1.9× for 500 and 2.6× for the 1,000 retweet threshold.

6.2 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 Poisson Pseudo-Maximum Likelihood (PPML) models (see Santos Silva and Tenreyro 2006), which naturally account for zeros in count data and handle heteroskedasticity without log-transforming the dependent variable, following Chen and J. Roth (2024).[14] We estimate:

$$E[Y_{i,t} \mid P_{i,t}] = \exp\left[\alpha + \beta P_{i,t} + \theta_1 X_{i,t} + \delta_{\tau(t)\times s(i)} + \gamma_{\tau(t)\times w(t)} + \eta_{h(t)} + \sum_{k\in K} \phi_k \, m_{ik}\right] \tag{1}$$

where $Y_{i,t}$ is the count of retweets for tweet i posted at time t. Because tweets are indexed at the hour level, we treat t as continuous and decompose it into year $\tau(t)$, calendar week $w(t)$, and hour of the day $h(t)$; the state $s(i)$ is a property of the tweet's author. α is the constant term. $\delta_{\tau(t)\times s(i)}$ is a year × state fixed effect.[15] $\gamma_{\tau(t)\times w(t)}$ is a year × calendar-week fixed effect that absorbs news and platform shocks specific to the time within a year. $\eta_{h(t)}$ is an hour-of-the-day fixed effect capturing diurnal cycles (systematic within-day patterns in Twitter engagement). $\sum_{k\in K} \phi_k \, m_{ik}$ are character fixed effects, where $m_{i,k} = 1$ indicates that character k appears in tweet i, so that identification comes from within-character variation in role assignment. $X_{i,t}$ is a vector of control variables at the tweet level, including the language metrics and measures of valence and emotions introduced in Subsection 5.2.[16] Robust standard errors (clustered at the week level here) handle over- and under-dispersion in the Poisson likelihood.

$P_{i,t}$ refers to our variable of interest. In our first main test, $P_{i,t}$ takes 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 $P_{i,t} = 0$

[14] As recommended, Subsection E.6 also shows results with OLS and a log transformation for the intensive and extensive margin. The Poisson variance-equals-mean assumption is not tested as it is not required for the consistency of the estimator (Wooldridge 1999).

[15] 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.

[16] There might also be measurement error due to misclassifications of the role labels $P_{i,t}$. If we assume that this is approximately mean-zero and independent of $Y_{i,t}$ conditional on covariates, it should act like classical measurement error, attenuating estimates toward zero under a linear approximation. However, LLM misclassifications could instead be correlated with virality, biasing our estimates upward. In our 500-tweet validation sample, we find no such correlation between LLM–human agreement and retweet counts (see Figure B.1).

are also on the same topic and contain at least one character. We use robust standard errors clustered at the week-level, mitigating potential over- and under-dispersion in the Poisson likelihood. β is our coefficient of interest and captures a semi-elasticity: $\exp(\beta) - 1$ gives the percentage change in expected retweets associated with political narratives. Accordingly the virality premium is proportional: a given percentage change implies more additional retweets for tweets that would have received more retweets absent the narrative. This is consistent with the log-log distribution in Figure 6, where the gap is largest at the most viral end.

Our regressions identify the conditional virality premium of narrative structure: how many more retweets a tweet receives on average when it contains a drama-triangle role, compared to a counterfactual neutral tweet. The counterfactual here should be thought of as a tweet posted at a similar time, discussing the same characters, from an author with similar characteristics — notably a similar number of followers — and using similar emotional and language features. Our approach therefore does not aim to explain virality as a dynamic phenomenon, which would require modeling how narrative use itself shapes an author's follower count over time, or how the ranking algorithm differentially amplifies certain content. For instance, the number of followers at the moment of posting can itself be partly shaped by past narrative use; we take it as given.

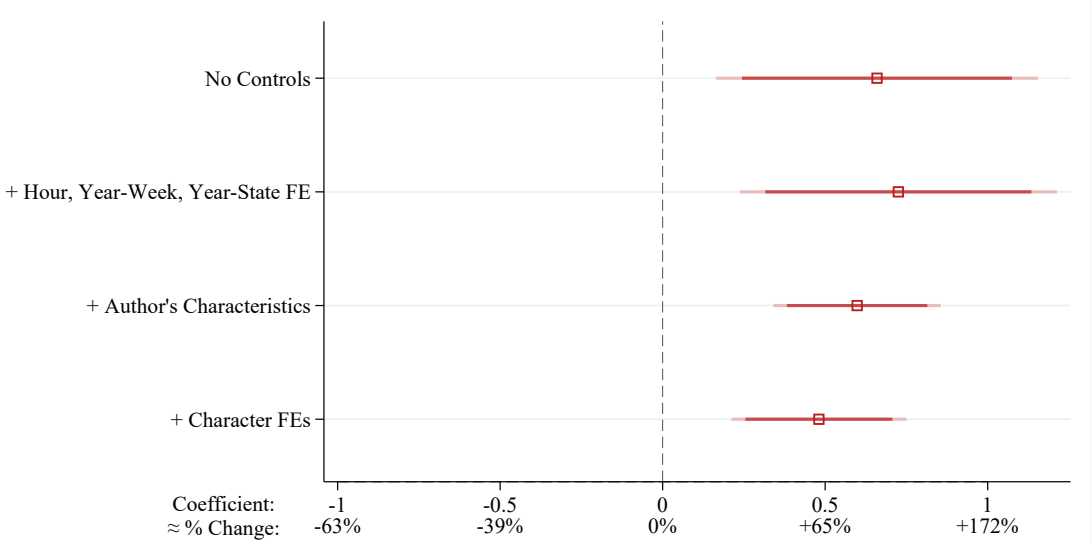
Figure 7 displays a comprehensive coefficient plot with six estimates ranging from the most simple to the most restrictive specification. The first row shows the simple correlation without any controls. Political narratives seem to be clearly more viral. The unconditional coefficient is large, positive, 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 concern 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 year $\times$ calendar-week fixed effects absorb news and 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 fixed effects capture diurnal cycles, referring to systematic within-day patterns in role use and virality on Twitter. Finally, year $\times$ state fixed effects allow both for heterogeneous effects of temporal shocks across states, and absorb 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%.

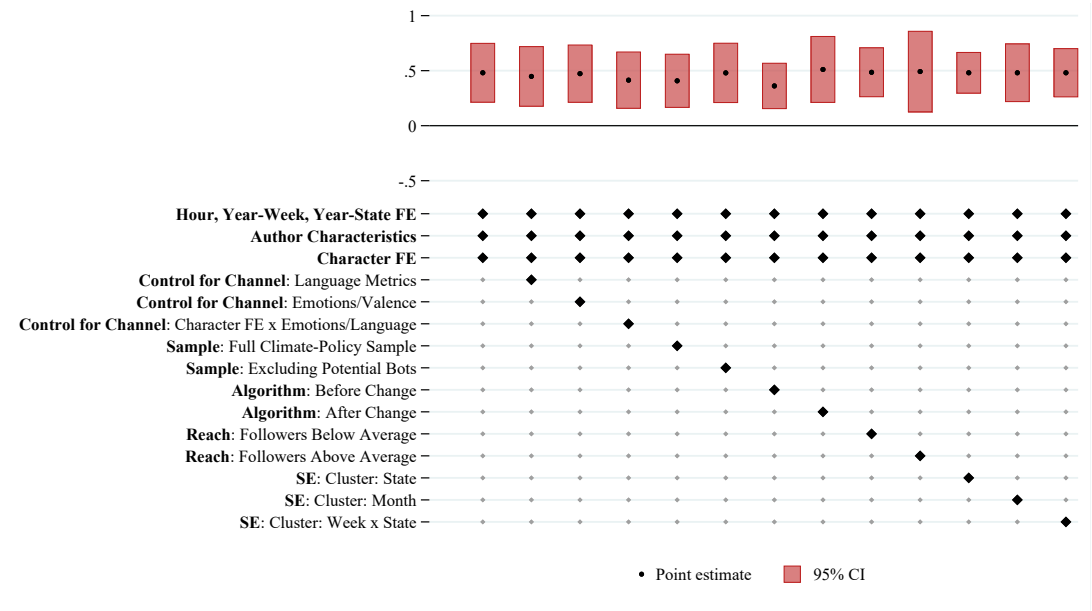
A second concern relates to author characteristics, which might correlate with both the usage of political narratives and virality. The most obvious difference is the number of followers: an account with many followers can go viral much more easily than a small account. If large accounts also tend to use more political narrative elements, this could lead to a spurious positive correlation between *political narrative* and retweets. By controlling for all available author characteristics, we only evaluate political narratives against neutral tweets from comparable authors. Controlling for author characteristics attenuates the implied virality premium from approximately 103% to 83%,

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

(a) ***Virality of Political Narratives vs. Neutral***



(b) ***Summary of Main Robustness Checks***



Notes: The figures report the main results of our analysis on the virality of political narratives. Panel 7a 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: $\approx e^{\beta} - 1$. The panels display results from increasingly restrictive models; the specification of the last model, is the one we use in the subsequent analysis. Author characteristics include: verified status, number of followers/followings, total tweets created, party affiliation, religiosity, higher education, and parenthood. The Full Climate-Policy specification does not restrict to tweets with predefined characters, expanding the baseline to all climate-policy tweets (including 16% with no character). Table D.1 shows a table with the full regression results. Panel 7b shows a summary of all main robustness checks we compute. Table D.2 displays the underlying coefficients.

while remaining clearly statistically significant. [17]

A third concern is that certain characters could be both more likely to be cast in drama triangle roles and generally tend to 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. 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."

6.2.1 Robustness Checks and Additional Results

Figure 7b summarizes our main robustness tests alongside the baseline estimate. We begin by controlling for emotions, language metrics, as well as for character FE interacted with net valence, anger, trust, reading ease, and lexical density. Authors framing Corporations as villain typically choose stronger verbs, moral language, and emotional words, which in turn affect virality. Including these features can reduce confounding from algorithmic surfacing but comes at the cost of conditioning on mediators (bad controls), so the main coefficient remains our preferred estimate.[18] Adding these controls only marginally reduces the effect size, indicating that the structured information we derive (character-roles and their combinations) goes beyond language metrics, emotions, or valence.

Our baseline sample $\mathcal{D}^{\text{rel}}$ excludes the roughly 16% of climate-policy tweets containing none of our ten predefined characters, so the premium could be an artifact of sample restriction. If these tweets were systematically less viral – as would be the case, e.g., for tweets about diffuse climate impacts without identifiable actors – their exclusion would *bias against* the estimated premium by removing low-engagement observations from the $P_{i,t} = 0$ comparison group. We therefore estimate a robustness specification on the full sample of climate-policy tweets, with those no-character tweets added to the comparison group. The coefficient remains positive and statistically significant, consistent with the *bias against* direction and indicating that the virality premium survives expanding the comparison group.

Bots could retweet political narratives disproportionately on social media, in which case the observed premium would partly reflect automated amplification rather than genuine human sharing. There is no perfectly reliable way to identify bots, so we follow the computer science literature and proxy for bot origin based on high posting volume, low engagement, and frequently duplicate text-posting. Dropping all suspicious tweets under these criteria is conservative: it removes tweets that may or may not be bots, biasing our test toward rejecting the main result. Both size and significance of the coefficient are barely affected, indicating that this is not driving our estimate.

The platform's algorithm could also shape our estimate, biasing the coefficient downwards

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

[18] Subsection E.1 provides a simple causal graph clarifying when language and emotions act as mediators versus confounders.

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 this sensitivity is to examine whether important changes in the algorithm affect the result. We therefore split the sample around the introduction of an algorithmically ranked timeline in February 2016, arguably the most important change in the Twitter algorithm during our sample period. The pre- and post-2016 coefficients are essentially indistinguishable, and interaction models in Subsection E.3 yield the same conclusion. Together with the earlier results on language and emotion controls, this reassures us that the narrative virality premium is not a mechanical result of algorithmic timeline curation.

The narrative premium might also differ across user types. 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. The below- and above-average reach coefficients show that the virality premium is present in both tiers, with slightly larger magnitudes for above-average reach accounts. Further checks shows that the results remain statistically significant at the 1%-level under three alternative standard-error clustering levels: state, month, and state $\times$ week. Subsection E.7 and Subsection E.8 show very similar results when using likes as an alternative outcome, and some interesting subtle differences when looking at replies. Subsection E.6 shows OLS specifications with log and asinh transformations, including extensive and intensive margin decompositions.

6.2.2 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 2 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 up to one character-role 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 157% 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 2 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

Table 2: ***Regression Results-Impact of Individual Roles on Virality***

Dependent Variable	**Retweets' Count**						
	Coeff./SE/p-value						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hero	0.410			0.401		-0.536	0.233
	(0.193)			(0.214)		(0.216)	(0.280)
	[0.034]			[0.060]		[0.013]	[0.406]
Villain		0.944		0.938	0.536		0.769
		(0.140)		(0.201)	(0.216)		(0.293)
		[0.000]		[0.000]	[0.013]		[0.009]
Victim			0.090	0.169	-0.233	-0.769	
			(0.326)	(0.289)	(0.280)	(0.293)	
			[0.782]	[0.559]	[0.406]	[0.009]	
Neutral					-0.401	-0.938	-0.169
					(0.214)	(0.201)	(0.289)
					[0.060]	[0.000]	[0.559]
Sample: Neutral	✓	✓	✓	✓	✓	✓	✓
Sample: Hero	✓			✓	✓	✓	✓
Sample: Villain		✓		✓	✓	✓	✓
Sample: Victim			✓	✓	✓	✓	✓
Percentage Change	50.7	157	9.4				
Mean Outcome (Control Group)	1.96	1.96	1.96	1.96	4.35	3.71	2.42
Pseudo R2	0.79	0.70	0.68	0.71	0.71	0.71	0.71
Observations	77,207	90,230	43,972	138,030	138,030	138,030	138,030

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, year × calendar-week, and year × state fixed effects, author characteristics, as well as character fixed effects. Standard errors are clustered at the week level.

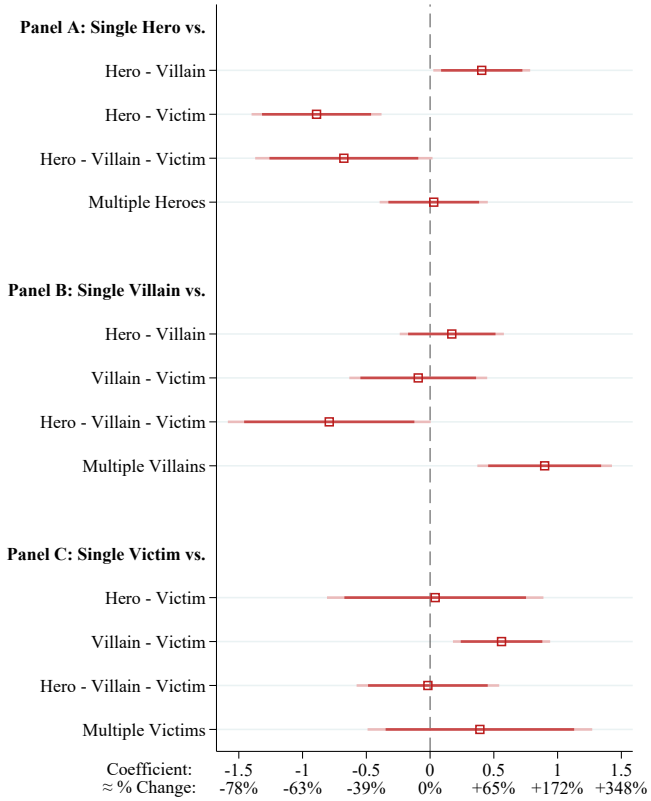
associated with a statistically significant virality advantage in any specification. While these are important insights for tweets with one role, we continue in the next section by shifting attention to tweets with multiple character-role combinations and differentiate between human and instrument character types.

6.2.3 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 C.4 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

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, year $\times$ calendar-week, and year $\times$ state fixed effects. Standard errors are clustered at the week level, covering all weeks in the study period. The underlying table is Table D.3.

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.

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. Such grander narratives may be better suited to media with fewer space constraints, yet they still act as the meta-narratives that simpler viral fragments draw on.

The results of adding additional roles to single hero, villain or victim narratives in Figure 8 reveal three main patterns. First, an additional villain character is consistently associated with further boosts to 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.

6.2.4 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 C.4 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.

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 93.7%. 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 31.0%. 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 (139.2%).

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.

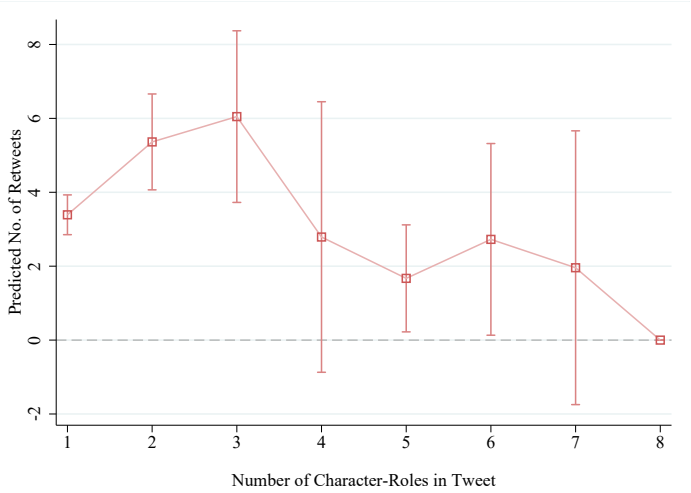
Figure 9b complements column 4 by plotting the predicted effects associated with additional character-roles. It turns out that adding up to 3 CRs boosts virality, but going to four and beyond leads to a clear decrease and partly insignificant point estimate. Combined with the prior results

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

(a) ***Heterogeneous Effects on Virality***

Dependent Variable	**Retweets Count**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Human	0.661	0.527		
	(0.137)	(0.186)		
	[0.000]	[0.005]		
Instrument	0.270	0.065		
	(0.238)	(0.206)		
	[0.255]	[0.751]		
Human × Instrument		0.280		
		(0.293)		
		[0.339]		
Sum of CRs			0.199	0.676
			(0.057)	(0.156)
			[0.001]	[0.000]
Sum of CRs Squared				-0.134
				(0.047)
				[0.005]
Mean Outcome Reference Group	2.176	2.176	2.176	2.176
Pseudo R2	0.63	0.63	0.63	0.63
Obs.	310,016	310,016	310,016	310,016

(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, year × calendar-week, and year × state fixed effects. Standard errors are clustered at the week level, covering the full time frame. 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. Marginal changes are reported in Table D.4.

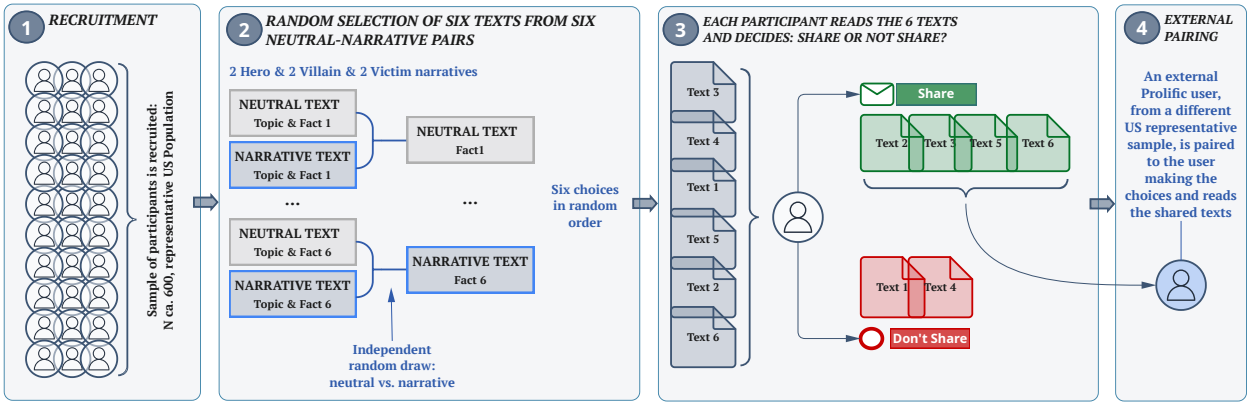
on role combinations, this effect has to be driven by multiple villains, combinations of heroes and villains and of villains and victims. At least suggestively, these patterns tentatively suggest certain strategies to boost virality: a stronger reliance on human characters, on villain characters, potentially combined with a hero or victim, but overall not more than three CRs.

6.3 Experimental Results: Virality

To address the most important of the limitations of the observational analysis, we conduct a pre-registered virality experiment for which we recruited a US-representative sample of 600 participants via *Prolific*. As depicted in Figure 10, each participant makes six real stakes sharing decisions in random order: each are independently and randomly assigned to see either the neutral or the narrative version and decide whether to share it with another Prolific participant in another separate experiment. Both versions contain identical factual information and differ only in whether characters are cast in drama triangle roles. The six text pairs cover all three roles (two hero pairs, two villain pairs, two victim pairs), span topics outside the observational setting – including organ donation,

national parks, health insurance, opioids, parking enforcement, and social media. Because each participant makes six binary decisions, the data form a within-subject panel.

Figure 10: ***Visualization of the Experimental Design: Virality Experiment***



Notes: The diagram illustrates the design of our virality experiments.

Table 3 reports linear probability model estimates of the treatment effect on the decision to share, with standard errors clustered at the participant level. Column 1 shows that exposure to the narrative version increases the probability of sharing by 14.9 percentage points ($p < 0.001$), relative to a control-group mean of 39%. Column 2 adds participant fixed effects to absorb all time-invariant individual characteristics; the estimate remains virtually unchanged at 15.7 percentage points. Columns 3–5 disaggregate by role type. Hero and villain narratives produce large and highly significant effects of 21.8 and 22.2 percentage points, respectively (both $p < 0.001$), each relative to a control-group mean of approximately 45%. The victim narrative effect is smaller at 4.6 percentage points (and still much smaller relative to the slightly lower control-group mean) and not statistically significant ($p = 0.165$). This suggests that the prior result that victim roles produce a weaker virality effect than hero or villain roles was not solely an artifact of the topic or social media setting.[19]

We also address two additional concerns about the generalizability of our observational results. First, Twitter may be a uniquely partisan platform which could mean that narrative effects on virality partly reflect in-group political dynamics rather than a more general mechanism. Second, the platform's design favoring short, high-impact communication may make narratives a disproportionally effective tool. We tackle both concerns using the experimental data. Appendix Table J.1 and Table J.2 subset participants by self-reported political leaning and by preferred platform of use (Twitter, Facebook, Instagram, or all three), respectively. The results consistently show that the narrative premium on the sharing decision is independent of political leaning and of the platform participants primarily use. Appendix Table J.3 further shows that social-media use intensity does not moderate the effect: the sharing premium is present for participants who use social media as little as 30 minutes per day and for those who use it three or more hours daily.

[19]Details, pre-registration link, and the full set of experimental texts are provided in Subsection H.1.

These experimental results further suggests that the positive association between political narratives and virality reflects a causal relationship. The within-subject design with random assignment directly addresses concerns about algorithmic amplification, bot activity, and author-level selection that are inherent to the observational setting. Because the experimental texts cover topics entirely outside the observational setting, the results demonstrate that the narrative virality premium generalizes beyond climate policy. The pattern across roles mirrors the observational findings: hero and villain narratives generate strong effects on sharing, while victim narratives show a weaker and less precisely estimated effect. The remaining limitation is that we cannot distinguish whether the additional virality premium of villain over hero roles was specific to social media or to the climate change policy topic. Nonetheless, the observational and experimental evidence clearly shows how our narrative framework helps to explain virality in a more systematic and structured way.

Table 3: ***Experimental Evidence on the Virality of Political Narratives: Panel Dataset at Prolific Participant Level***

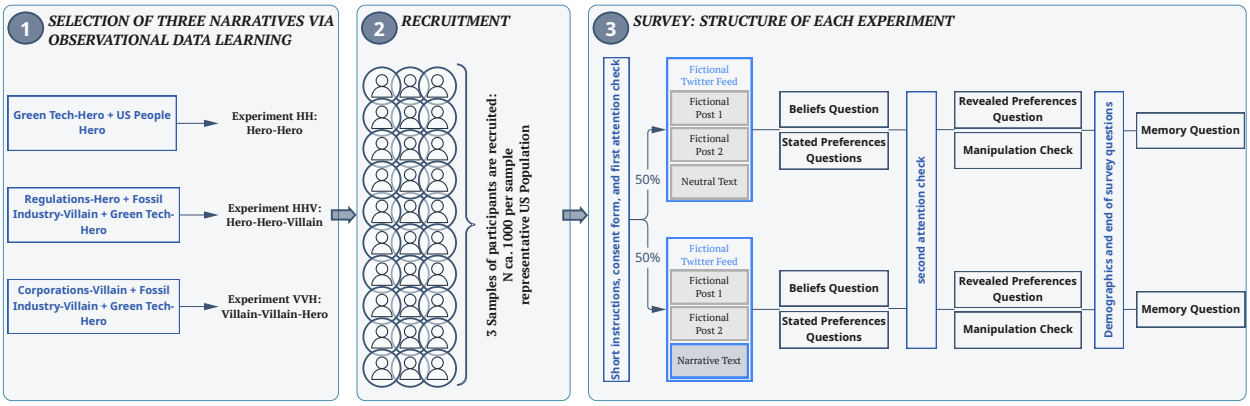
Dependent Variable	**Decision to Share Content**				
	All Types of Narratives		**Hero Narratives**	**Villain Narratives**	**Victim Narratives**
	Coeff./SE/p-value				
	(1)	(2)	(3)	(4)	(5)
Treatment	0.149	0.157	0.218	0.222	0.046
	(0.018)	(0.018)	(0.035)	(0.036)	(0.033)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.165]
Participant FE		✓	✓	✓	✓
Mean Outcome Control Group	0.39	0.39	0.45	0.45	0.27
Observations	3,469	3,468	1,156	1,156	1,156

Notes: The table reports OLS estimates from models estimating the impact of narratives on virality. The data is collected through our first experiment, where 600 participants were asked to read six short text extracts and choose whether they would share the text with another Prolific participant. The pairing with the receiving participant is real and the choice is a real stake outcome. For each choice participants were randomly assigned to see either the neutral or the narrative version of the same text. Neutral and narrative contain the same factual information and differ only in the presence of narrative roles. The data is organized in a panel where there are six entries per participant; the choices were in random order and we use participant FE to exploit within participant variation. The six pairs were divided in: two pairs focused on hero narratives, two on villains and two on victims. Columns 1-4 report results for the full sample of participants while columns 5-7 subset participants by their self reported political stance. Table K.1 shows the results using randomization inference for p-values.

7 Experimental Results: Persuasiveness 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 narratives on average go more viral, with villain characters being even more efficient than hero characters. Repeated exposure to media content is well-documented to shift attitudes and behavior at population scale (DellaVigna and Gentzkow 2010), but is there already an effect on beliefs, preferences and memory from single exposure to narratives? To test this, we

Figure 11: ***Visualization of the Experimental Design: Persuasion Experiments***



Notes: The diagram illustrates a visualization of the design of our persuasion and memory experiments.

run three parallel survey experiments, each testing climate-policy narratives combining Green Tech cast as hero with different character-roles, building on combinations that were also frequent and viral in our observational data.

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.

This section is limited in scope to climate change policy as a topic, replicating across other topics and role configurations constitutes an important avenue for future work. The remainder of the section is organized as follows. First we provide an overview on the design of our experiments in Subsection 7.1, second we show our results on beliefs and preferences in Subsection 7.2, and thirdly, we explore the impact of political narratives on memory in Subsection 7.3.

7.1 Experimental Design

We conduct three pre-registered online experiments, each following an identical design (see Figure 11. 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. Subsection H.3 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 as a reminder, but wit the relevant third post blurred (full questionnaires in Subsection H.2).

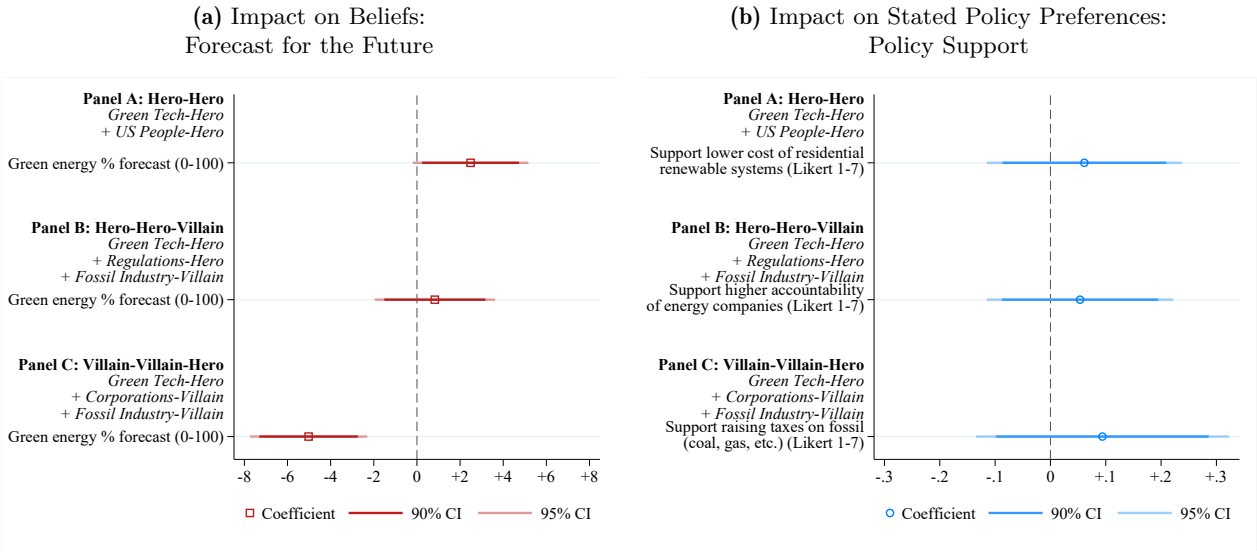
We use a combination of different types of questions, explained in more detail in the respective following sections. Our belief question asks participants to predict the share of renewable energy in the US by 2035. As this is unknown, there is no incentive to correctly guess; however, we also see no 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.

7.2 Experimental Results: Virality, Beliefs, and Preferences

We turn to the effects of the three political narratives on beliefs and stated preferences. Figure 12a 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

Figure 12: ***Experimental Results-Impact of Political Narratives on Beliefs and Policy Preferences***



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 12a, we show the impact of the narrative treatment on beliefs, namely 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.'*. Table I.1 and Table L.4 show the corresponding regression models, with controls, and using randomization inference for p-values, respectively. Figure L.2a shows the results without controls. In Figure 12b, 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 I.2 and Table L.5 show the corresponding regression models, with controls, and using randomization inference for p-values, respectively. Figure L.2b shows the results without controls.

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. The three experiments trace a monotonic gradient in belief effects, from positive under HH through near-zero under HHV to negative under VVH, consistent with the logical implication that as the share of villain roles in the treatment increases, Green Tech's success becomes less plausible. 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 12b shows effects on stated policy preferences, using the same vertical structure as for beliefs. To do so, we measure support for specific climate

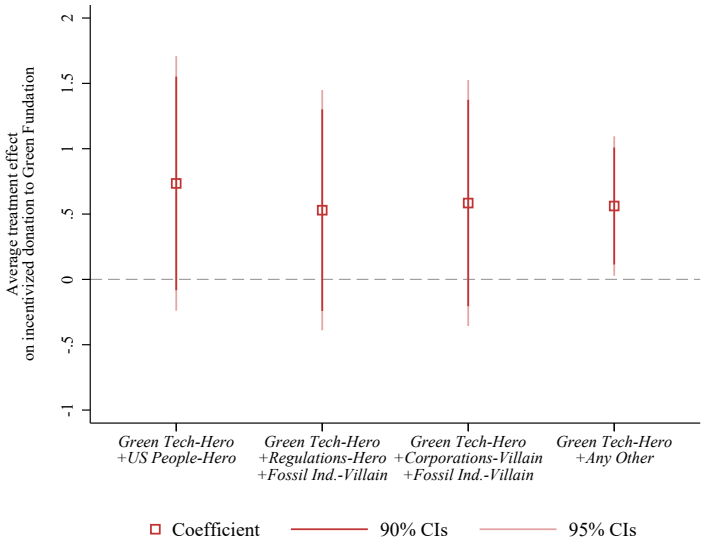
change policy proposals that fit the respective character combinations in the treatment-control pairs. For the Hero-Hero combination, the question is about support for higher government subsidies for residential renewable energy systems. For the Hero-Hero-Villain combination, the policy proposal is about 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, despite our design being adequately powered to detect relatively small effects (see Subsection L.2). 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).[20]

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 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 13 displays the coefficients together with 90% and 95% confidence intervals from each individual experiment, along with the pooled estimate for all experiments. The point estimates are of similar magnitude and point in the same direction across the three experiments, suggesting a robust treatment effect of narratives despite limited power in any individual study. 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.

[20] In Table F.3, 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 correlates negatively with support for renewable energy, and that the frequency of Regulation-Villain correlates negatively with support for the Environmental Protection agency.

Figure 13: ***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 I.3 and Table L.6 show the full corresponding regression models, with controls and also when using randomization inference for p-values. Figure L.3 shows the results without controls.

7.3 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, C. 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. [21]

Figure 14a shows the results of the exposure to political narratives on memory, pooling all three treatment-control pairs. The top panel shows the coefficient measuring the effect of being exposed to the narrative treatment on remembering the factual information contained in the tweet, the bottom 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 L.2 and Table L.3). 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.

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 14b 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.[22] 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.

One plausible reason is that attention influences how people process and store information (Loewenstein and Wojtowicz 2025). The drama triangle roles reflect a salient category, e.g. a threat (Vuilleumier 2005), that resonate with people and their whole set of memories. The representation of characters in these roles plausibly has an inherent advantage for encoding information and storing it efficiently into memory. If a character in memory is linked to a larger semantic cluster, like an archetypal role that people are familiar with, it can be retrieved more easily.

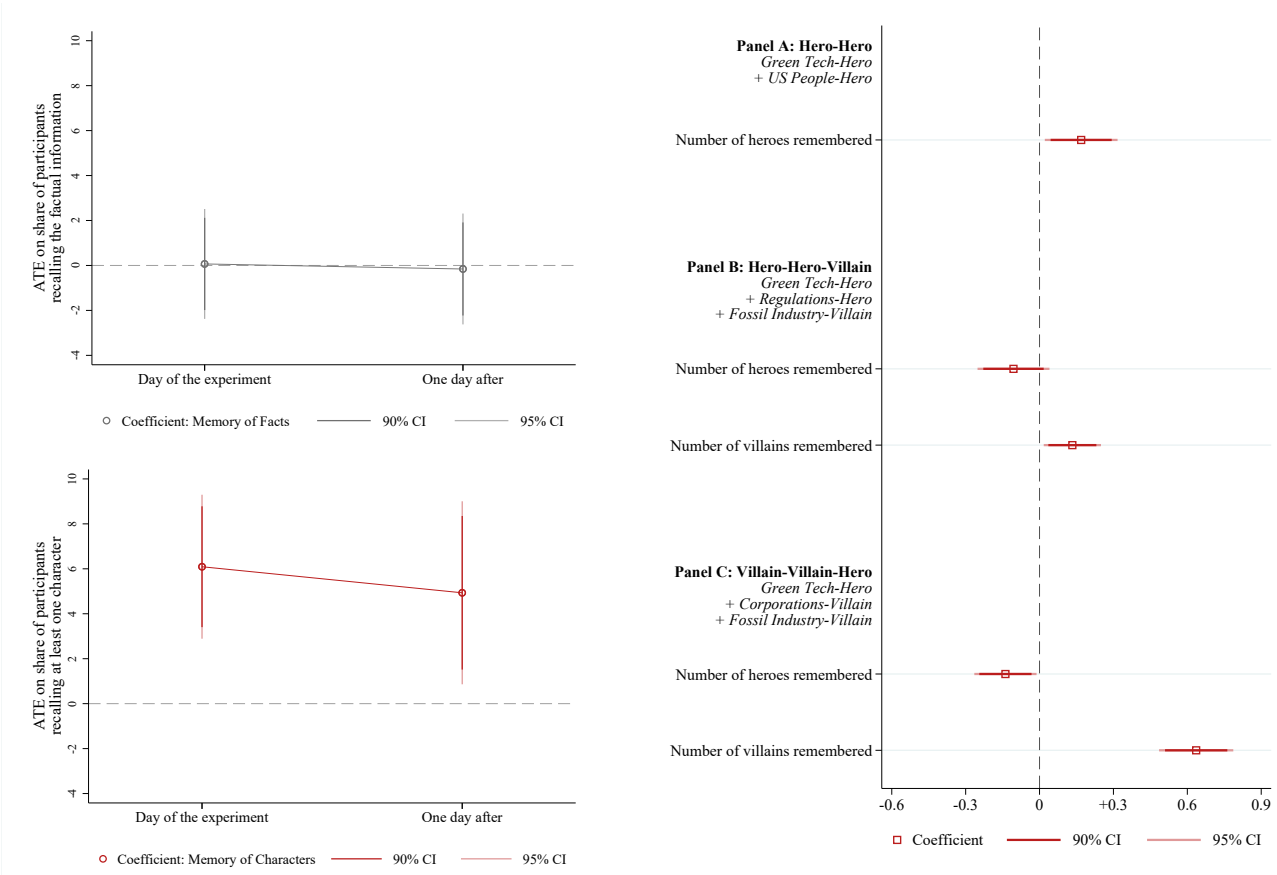
[21] The question posed on the day of the experiment serves as basis for the results of our manipulation check as well, in Table L.1.

[22] 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.

Figure 14: ***Experimental Results-The Impact of Political Narratives on Memory (Pooled Sample)***

(a) Memory: Recall of Factual Information and Character Presence

(b) Memory: Recall of Characters by Role



Notes: The figure reports OLS estimates of the treatment effect on participants' memory. The top plot in Figure 14a 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. The bottom plot in Figure 14a 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 I.4 and Table L.7 show the corresponding regression models, with controls and using randomization inference for p-values, respectively. Figure L.4a shows the results without controls. Figure 14b 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 I.5 and Table L.8 show the corresponding regression models, with controls and using randomization inference for p-values, respectively. Figure L.4b shows the results without controls.

A key insight in Graeber, C. 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.[23]

Pure emotion explanations implausible: While our approach cannot perfectly isolate the effect of role assignment from emotions or causal sequences, since both can serve to signal roles, our results indicate that emotions are not the sole driving force. The donation outcome provides the clearest evidence: Green Tech is cast as hero in all three experiments, and the donation effect is positive and similar in magnitude across conditions, despite valence differences that range from strongly positive (Hero-Hero) to strongly negative (Villain-Villain-Hero) (Figure L.5). If emotions alone drove the effect, the donation response should track valence; it does not. It also does not track anger patterns, highlighted as a key emotion in Algan et al. (2025). The memory results point in the same direction: character recall is similar in the two villain-heavy conditions despite different valence and anger levels. We cannot offer a definitive decomposition, but this evidence, combined with the observational results, strongly indicates that role assignment matters beyond emotional tone.

8 Conclusion

In conclusion, our study demonstrates that the political narrative framework provides a powerful lens for understanding how such 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 political narratives go viral. Political narratives, on average, receive about sixty per cent more retweets in expectation. 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 157 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.

[23] Our fact question is a verbatim numeric recall of a specific number, 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, C. 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, C. Roth, and Zimmermann (2024), who also document better recall of information type and direction, not specific numbers.

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 donate more to an organization aligned with the character cast as hero in the narrative. 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 a complementary dimension that pure sequence based approaches do not explicitly encode. They also reach 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. 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 costs have fallen substantially, the remaining tasks for new domains are careful character definitions and some human validation.

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.

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

Andre, Peter, Ingar Haaland, 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, Ruben Durante, et al. (2021). "Visual Representation and Stereotypes in News Media". In: *Center for Law & Economics Working Paper Series* 2021.15.

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

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.

Barron, Kai and Tilman Fries (2024). "Narrative Persuasion". In: *WZB Discussion Paper* 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.

Bénabou, Roland, Armin Falk, and Jean Tirole (2020). "Narratives, Imperatives, and Moral Reasoning". In: *NBER Working Paper* 24798.

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.

Brady, William J. et al. (2017). "Emotion shapes the diffusion of moralized content in social networks". In: *Proceedings of the National Academy of Sciences* 114.28, pp. 7313–7318.

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

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

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.

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.

DellaVigna, Stefano and Matthew Gentzkow (2010). "Persuasion: Empirical Evidence". In: *Annual Review of Economics* 2.1, pp. 643–669.

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, Ruben Durante, Elliot Motte, et al. (2024). "Experience, Narratives, and Climate Change Beliefs". In: *SSRN Electronic Journal*.

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

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.

Gehring, Kai, Joop Age Adema, and Panu Poutvaara (2022). "Immigrant Narratives". In: *CESifo Working Paper* 10026.

Gehring, Kai and Matteo Grigoletto (May 2023). "Analyzing Climate Change Policy Narratives with the Character-Role Narrative Framework". In: *CESifo Working Paper* 10429.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019). "Text as Data". In: *Journal of Economic Literature* 57.3, pp. 535–74.

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

Graham, Jesse, Jonathan Haidt, and Brian A. Nosek (2009). "Liberals and conservatives rely on different sets of moral foundations". In: *Journal of Personality and Social Psychology* 96.5, pp. 1029–1046.

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.

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 Working 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*.

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.

Müller, Karsten and Carlo Schwarz (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.

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

Salminen, Joni et al. (2020). "Developing an Online Hate Classifier for Multiple Social Media Platforms". In: *Human-centric Computing and Information Sciences* 10.1, p. 1.

Santos Silva, J. M. C. and Silvana Tenreyro (2006). "The Log of Gravity". In: *Review of Economics and Statistics* 88.4, pp. 641–658.

Serra-Garcia, Marta (2025). "The Attention–Information Tradeoff". In: *CESifo Working Paper* 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.

Vosoughi, Soroush, Deb Roy, and Sinan Aral (2018). "The spread of true and false news online". In: *Science* 359.6380, pp. 1146–1151.

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.

Wooldridge, Jeffrey M. (1999). "Distribution-free estimation of some nonlinear panel data models". In: *Journal of Econometrics* 90.1, pp. 77–97.

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[24]: 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, critiques, or

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

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 Data Sources and Variables

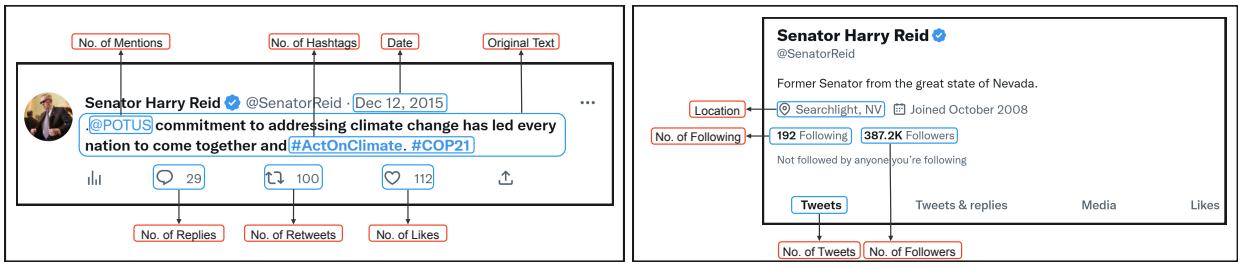
In the last section of this Appendix we provide additional information on the data sources and variables used in our analysis. We start with Table A.1 that lists the various sources of the data used. We dive in the most important source, the Twitter API, in Figure A.1. The figure offers a visual reference 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.

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.

Figure A.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 4 of the paper we describe our data.

Table A.2 and A.3 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 A.2 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 A.3 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.

Table A.2: ***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	$n \in [0; 24{,}080]$	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)	$n \in [0; 110]$	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	$n \in [0; 850{,}382]$	Twitter APIv2
No. of Tweets	Total tweets produced by the author up to posting	$n \in [0; 3{,}671{,}810]$	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 4 of the paper, we describe our data in more detail.

Table A.3: ***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 regarding the text metrics, valence, and emotions in text. In Section 4 of the paper we describe our data.

B Validation of Characters and Roles' Classification

B.1 LLM 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. Large 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 LLMs classification 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.

This appendix compares LLM's classification to that of human coders on the same tweets. Given that we think of LLM's output as reflecting an average representative human coder, this exercise explores how closely LLM's interpretations align with those of human coders. 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, an online platform where participants give their availability for short term tasks, and are paid based on time spent on these tasks. 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 exercise classifies 500 tweets randomly selected among those identified by the LLM 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. In total, 80 workers successfully completed the qualification test and were invited to participate; of these, 28 actually classified tweets. Workers were free to decide how many tweets to classify: some coded as few as one tweet, while others classified up to 130, with an average of 36 tweets per worker. We kept the task open until each of the 500 tweets had been coded by two different workers.

The classification task was divided into two phases, mirroring as closely as possible the structure of the prompt used for the LLM's classification. For each character, human coders were first asked whether the character was present in the tweet. For characters identified as present, coders then indicated whether the character was cast as a hero, villain, victim, or none of these roles. We used these classifications to compare human coders to each other and to the LLM's classification.

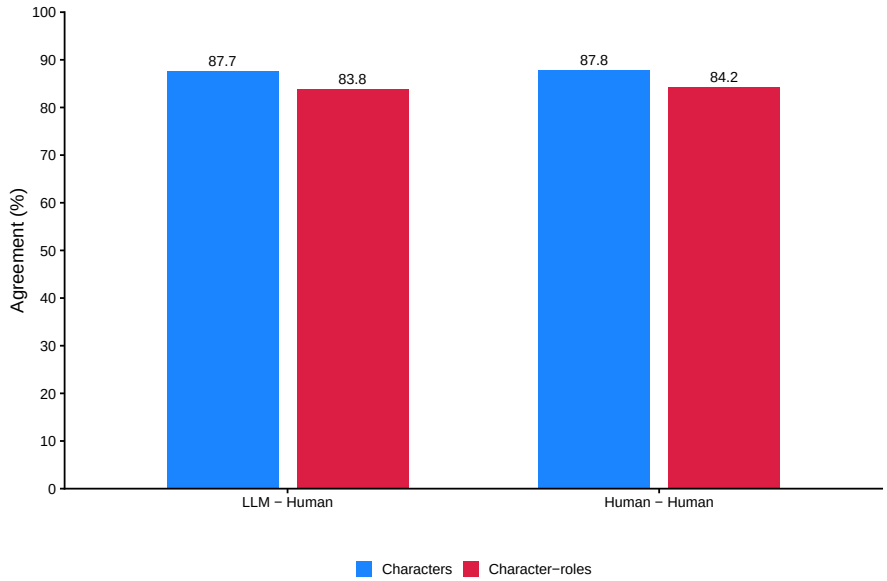
Comparison of the Classification of Characters and Character-Roles

In the first part of this analysis, we compare classification of characters and character-roles.

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 LLM and Human Coders, and between pairs of Human Coders. Figure B.1 summarizes the results. The figure shows the share of tweets where the LLM 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, the LLM 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 the LLM and a Human Coder) agreed that the character or character-role was absent.

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



Notes: The figure shows agreement rates between the LLM and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected relevant tweets. All tweets were classified by the LLM 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). The LLM's classification 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 the LLM and a Human Coder) agreed that a character or role was absent.

Overall, we find a high level of agreement between Human Coders and the LLM. On average, the LLM and Human Coders classified the presence of characters the same way in 87.7% of cases.

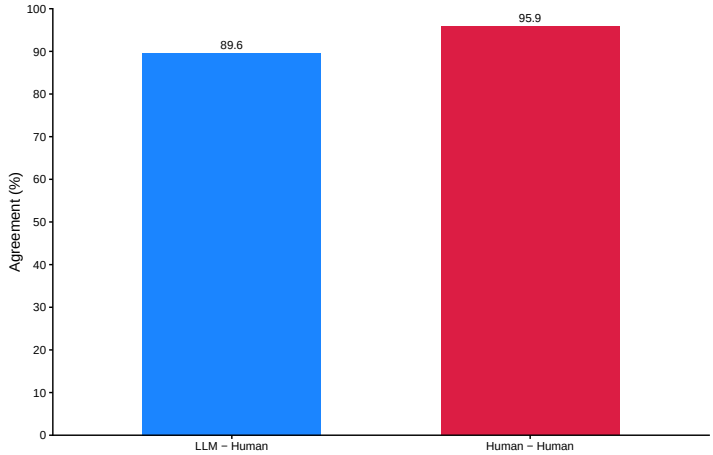
Agreement on character-role classifications was similarly high, at 83.8%. Notably, these rates closely mirror the agreement between Human Coders themselves. Two independent Human Coders agreed on the character in 87.8% of cases and on the character role in 84.2% of cases. These results support our view of the LLM as an average or representative Human Coder.

As explained above, our main agreement measures include 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 LLM-based coding.

In the second step, we examine agreement on role assignment conditional on character presence. The previous analysis (Figure B.1) captured agreement on both dimensions jointly – whether a character was present and which role, if any, was assigned. Here, we restrict to tweets where the two human coders agreed that a specific character was present, and compare only their role assignments. This isolates the harder question: given that a character appears in the tweet, do coders agree on whether it is cast as a hero, villain, or victim? Based on our argument that the drama triangle (Karpman 1968) reflects a natural way humans interpret reality, we expect high agreement at this stage.

Figure B.2 shows the results of this second exercise. On the left, the blue bar indicates agreement between the LLM 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 the LLM 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.

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



Notes: The figure shows agreement rates between LLM and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected tweets. All tweets were classified by the LLM 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 the LLM 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.

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 the LLM 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.

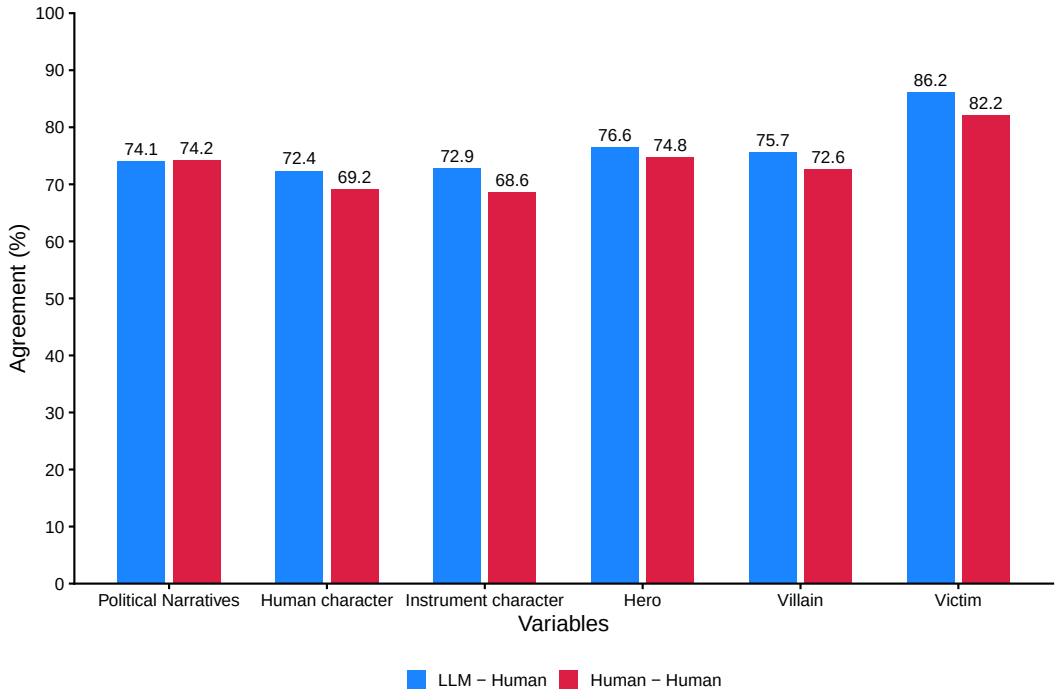
Political Narratives We compute the Political Narratives variable at the tweet level, defined as containing at least one character-role combination. We assess agreement rates between the LLM and Human Coders (blue) and between pairs of Human Coders (red). As shown in Figure B.3, agreement levels are very similar: around 74% for LLM-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, the LLM 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 the LLM 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 supports the validity of our approach. Figure B.3 shows that the LLM 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 the LLM 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.

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



Notes: The figure shows agreement rates between LLM and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected relevant tweets. All tweets were classified by the LLM as containing at least one of the 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 the LLM and the two humans. The red bar indicates the average level of agreement between two humans.

Misclassification and Virality

A natural concern is whether LLM misclassification is systematically related to virality. If the tweets the LLM gets wrong are precisely the ones that spread the most, this would introduce a bias in our analysis. To investigate this, we examine whether disagreement between the LLM and human coders on political narrative classification is correlated with retweet counts.

We construct the sample as follows. Among the 500 tweets in our validation exercise, we retain those for which the two human coders agreed on whether a political narrative was present or absent. This yields 371 tweets, which we split into two groups: those where the LLM agreed with the human consensus ($N = 306$) and those where it did not ($N = 65$).

We find no evidence that LLM misclassification is correlated with virality. Tweets where the LLM disagreed with human coders had, on average, 0.86 retweets, compared to 0.53 for tweets where it agreed. The difference of 0.34 retweets corresponds to 0.11 standard deviations, and a Welch two-sample t-test does not reject equality of means ($p = 0.60$). To further assess this, we regress $\log(1 + \text{retweets})$ on a disagreement indicator with robust standard errors; the estimated coefficient is 0.003 (SE = 0.077, $p = 0.97$), indicating no systematic relationship between misclassification and tweet virality. As a non-parametric check, the Spearman rank correlation between the disagreement indicator and log(1+retweets) is $\rho = -0.008$ ($p = 0.87$). Across all three tests, LLM misclassification at the narrative level is uncorrelated with retweet counts, suggesting that the LLM's errors are not systematically concentrated among more viral tweets.

B.2 LLM vs Alternative NLP Methods

To complement Subsection B.1, we compare the LLM's output to a rule-based NLP pipeline that assigns both characters and roles, from here on *NLP (Dictionary + Dependency Parsing)*. We apply the same validation framework to both methods, comparing their classifications to human coders along four dimensions: *(i)* character presence (including both positive and negative cases), *(ii)* exact character-role assignment (including absence), *(iii)* political narrative detection at the tweet level, and *(iv)* character-role assignment conditional on human agreement that a character is present. For each dimension, we report both raw agreement rates and standard classification metrics – accuracy, precision, recall, and F1 – using human coders as ground truth. This parallel framework allows us to directly compare the rule-based pipeline, the LLM-based classifier, and inter-human agreement under identical criteria.

Results are reported in Figure 2, described in the paper Section 4. Figure 2a shows agreement rates between pairs of human coders, between the LLM-based classifier and human coders, and between the *NLP (Dictionary + Dependency Parsing)* pipeline and human coders. Figure 2b reports accuracy, precision, recall, and F1 for each method. Additionally, Table B.1 provides the same performance measures by role.

Method

The *NLP (Dictionary + Dependency Parsing)* is a rule-based pipeline designed to identify climate-related characters in tweets and assign them narrative roles based on lexical cues and syntactic structure. For each character, we compile high-precision identifiers and higher-recall terms; we also define role dictionaries for hero, villain, victim, and neutral. Heroes are framed as advancing climate solutions or accountability (e.g., lead, champion, invest, sue); villains as obstructing or deceiving (e.g., block, undermine, deny science, greenwash, pollute, bribe); victims as harmed or burdened (e.g., suffer, devastated, displaced, vulnerable, frontline). Neutral is the default when a character is detected but no role cues co-occur. We report the full list of terms used in this pipeline below.

Next, we combine these dictionary matches with syntactic information to assign roles more precisely. Each tweet is parsed with spaCy to obtain sentence boundaries, token lemmas, and the dependency tree. Within each sentence, we *(i)* detect which character categories are mentioned using the character dictionaries and *(ii)* identify candidate role indicators using both lemma matching (to capture inflected verbs/adjectives) and phrase matching (to capture multi-word expressions). For every detected character mention, we then compute its proximity to each candidate role indicator in the same sentence using dependency-path distance (the shortest path through the dependency tree); when no role indicator is present in the character's sentence, we fall back to linear token distance across sentences. We assign a character the role associated with the closest role indicator in its sentence; a character whose tweet contains no role indicator anywhere remains Neutral. When a character is mentioned multiple times within a tweet, the pipeline preserves first-mention priority when assigning the character's primary role.

Table B.1: ***Performance by Narrative Role (Villain, Hero, Victim)***

Role	Comparison with Human Coders	Accuracy	F1	Precision	Recall
Villain	LLM	89.9	87.7	88.9	86.6
	Dictionary + Dependency Parsing	60.5	15.8	70.0	8.9
Hero	LLM	87.5	82.1	88.5	76.6
	Dictionary + Dependency Parsing	65.0	37.1	56.5	27.7
Victim	LLM	94.7	47.4	100.0	31.0
	Dictionary + Dependency Parsing	90.7	5.4	12.5	3.4

Notes: This table disaggregates Panel B of Figure 2 by narrative role. The sample is restricted to character-level observations where both human coders agreed on the exact role assignment (Villain, Hero, or Victim). For each role, we apply a one-vs-rest evaluation: the listed role is the positive class, and any other prediction (including Neutral or absent) counts as negative. Accuracy is defined as $(TP+TN)/(TP+TN+FP+FN)$, precision as $TP/(TP+FP)$, recall as $TP/(TP+FN)$, and $F1 = 2 \cdot \text{precision} \cdot \text{recall}/(\text{precision} + \text{recall})$.

Terms Used in the NLP Pipeline

Character Dictionaries:

Developing Economies developing countries, developing world, emerging economies, emerging markets, global south, global-south, low-income countries, middle income countries, low and middle income countries, lmic, lmics, least developed countries, ldc, ldcs, climate vulnerable countries, vulnerable nations, frontline nations, g77, g-77, group of 77, g77+china, small island developing states, sids, small island states, island nations, climate finance, adaptation finance, mitigation finance, finance for adaptation, loss and damage, technology transfer, capacity building, green climate fund, gcf, adaptation fund, gef, global environment facility, clean development mechanism, cdm, sub-saharan africa, east africa, west africa, south asia, southeast asia, latin america, central america, the caribbean, caribbean nations, andean, bangladesh, pakistan, india, indonesia, philippines, vietnam, nigeria, kenya, ethiopia, tanzania, uganda, haiti, honduras, guatemala, maldives, tuvalu, vanuatu, fiji, brics, south nations, poor countries, poorer countries, africa, island countries, climate reparations, climate aid, foreign aid, south-north, north-south, climate debt, historical responsibility.

US Democrats democrat, democrats, democratic party, dems, dnc, house democrats, senate democrats, blue states, dccc, dscc, liberal, liberals, progressive, progressives, center-left, the left, left-wing, leftwing, #democrats, #voteblue, #bluewave, #resist, obama, barack obama, michelle obama, biden, joe biden, kamala harris, harris, clinton, hillary clinton, bill clinton, pelosi, nancy pelosi, schumer, chuck schumer, bernie, bernie sanders, aoc, alexandria ocasio-cortez, elizabeth warren, warren, john kerry, kerry, al gore, gore, jay inslee, inslee, ed markey, markey, katie porter, barbara boxer, lib, libs, leftist, leftists, socialist, socialists, dem governor, dem caucus.

US Republicans republican, republicans, gop, grand old party, rnc, house republicans, senate republicans, red states, freedom caucus, tea party, conservative, conservatives, right-wing, rightwing, the right, #gop, #maga, #tcot, #trump2020, trump, donald trump, mike pence, pence, mitch mcconnell, mcconnell, paul ryan, ryan, john boehner, boehner, ted cruz, cruz, marco rubio, rubio, mitt romney, romney, newt gingrich, gingrich, sarah palin, palin, rick perry, perry, scott pruitt, pruitt, andrew wheeler, wheeler, jim inhofe, inhofe, john barrasso, barrasso, kevin mccarthy, mccarthy, rino, alt-right, congressional gop, gop leadership, gop senators, gop house.

Corporations corporation, corporations, corporate, corporate america, multinational, conglomerate, fortune 500, wall street, big business, boardroom, shareholders, ceo, c-suite, executives, share buyback, stock buybacks, dividends, quarterly earnings, fiduciary duty, lobbyist, lobbyists, corporate lobby, lobbying, dark money, super pac, pac money, campaign donations, corporate donations, regulatory capture, revolving door, u.s. chamber of commerce, national association of manufacturers, nam, business roundtable, trade association, industry group, special interests, corporate welfare, bailout, subsidies, big corp, corporate greed, greedy corporations, mega-corp, money in politics.

US People people, the public, citizens, voters, taxpayers, ratepayers, consumers, families, working families, working class, middle class, workers, blue collar, communities, local communities,

residents, neighbors, homeowners, renters, kids, children, future generations, grandkids, next generation, farmers, ranchers, rural communities, coastal communities, low-income families, poor families, americans, ordinary americans, everyday people, the masses, folks, hardworking americans.

Emission Pricing carbon tax, revenue-neutral carbon tax, carbon fee, carbon fee and dividend, fee and dividend, tax-and-dividend, carbon dividend, price on carbon, carbon price, carbon pricing, cap and trade, cap-and-trade, emissions trading, emissions trading scheme, ets, carbon market, carbon markets, allowance trading, permit trading, auction allowances, permit auction, carbon credits, offsets, carbon offsets, voluntary carbon market, border carbon adjustment, carbon border adjustment, bca, rggi, regional greenhouse gas initiative, california cap-and-trade, citizens climate lobby, ccl, climate leadership council, carbon permit, carbon allowances, internal carbon price, social cost of carbon, scc.

Regulations epa, environmental protection agency, clean air act, nepa, endangerment finding, rulemaking, regulatory, regulation, regulations, compliance, enforcement, permit, permits, permitting, clean power plan, cpp, new source performance standards, nsps, vehicle emissions standards, tailpipe standards, fuel economy standards, cafe standards, nhtsa, methane rule, methane regulations, ozone standards, so2 limits, nox limits, renewable portfolio standard, rps, clean energy standard, clean electricity standard, ces, zero emission vehicle mandate, zev mandate, zev, building codes, energy code, efficiency standards, ban on fracking, fracking ban, drilling ban, moratorium on drilling, pipeline permit, pipeline permitting, ban, bans, mandate, mandates, standard, standards, rule, rules, roll back regulations, deregulation, environmental regulation, climate law, climate laws.

Fossil Industry fossil fuel industry, fossil fuels industry, oil and gas industry, oil industry, gas industry, coal industry, petroleum industry, big oil, big coal, oil lobby, coal lobby, gas lobby, oil drilling, offshore drilling, arctic drilling, fracking, hydraulic fracturing, shale drilling, tar sands, oil sands, bitumen, lng exports, liquefied natural gas, pipeline company, oil pipeline, gas pipeline, refinery, refineries, keystone xl, dakota access pipeline, dapl, line 3, line 5, permian basin, bakken, eagle ford, marcellus, utica shale, american petroleum institute, api, opec, exxon, exxonmobil, chevron, bp, shell, conocophillips, phillips 66, valero, marathon petroleum, halliburton, schlumberger, enbridge, kinder morgan, energy transfer, koch industries, koch brothers, peabody, arch coal, oil, coal, natural gas, gasoline, diesel, petroleum, drill, drilling, pipeline, frack, frackers, fracked.

Green Technologies renewable energy, renewables, clean energy, clean power, solar power, solar panels, rooftop solar, solar farm, community solar, wind power, wind turbines, wind farm, offshore wind, onshore wind, geothermal energy, geothermal, hydropower, hydroelectric, run-of-river, electric vehicle, electric vehicles, ev charging, ev charger, charging station, battery storage, grid storage, energy storage, lithium-ion, heat pump, building electrification, electrification, smart grid, microgrid, demand response, green hydrogen, electrolyzer, electrolyzers, energy

efficiency, weatherization, home insulation, led lighting, tesla, powerwall, solarcity, sunrun, first solar, enphase, solaredge, solar, wind, renewable, ev, evs, battery, batteries, hydrogen, efficiency, electrify, electrified.

Nuclear Energy nuclear power, nuclear energy, nuclear plant, nuclear reactor, reactor core, advanced nuclear, generation iv, gen iv, small modular reactor, smr, smrs, pressurized water reactor, pwr, boiling water reactor, bwr, molten salt reactor, msr, fast reactor, breeder reactor, thorium reactor, thorium, uranium, enrichment, fuel rods, spent nuclear fuel, dry cask storage, nuclear waste, radioactive waste, high-level waste, yucca mountain, nrc, nuclear regulatory commission, meltdown, radiation, containment, chernobyl, fukushima, three mile island, vogtle, diablo canyon, indian point, palo verde, westinghouse, nuscale, nuclear, reactor.

Role Dictionaries:

Villain block, obstruct, stonewall, stall, delay, derail, sabotage, gut, weaken, undermine, defund, dismantle, roll back, rollback, deregulate, loophole, exempt, deny, denialist, deny science, reject science, lie, mislead, deceive, gaslight, cover up, disinformation, misinformation, propaganda, greenwash, greenwashing, astroturf, astroturfing, pollute, poison, contaminate, spew, dump, spill, leak, destroy, wreck, ruin, ravage, bribe, bribery, bought off, sellout, payoff, corrupt, crooked, captured, greedy, criminal, fraud, scam, attack, blame, betray, sell out, rig, rigged, con, profit, cash in, exploit, extort, gouge, steal, rob, evil, toxic, dirty, anti-science, denier, climate hoax, puppet, shill, cronies.

Hero lead, champion, push, advance, pass, enact, implement, strengthen, restore, expand, fight, stand up, defend, protect, save, help, support, invest, fund, build, deploy, scale, innovate, accelerate, transition, decarbonize, phase out, hold accountable, sue, lawsuit, prosecute, investigate, bold, ambitious, historic, transformative, clean, sustainable, responsible, visionary, solution, act, action, deliver, get it done, breakthrough, progress, win, victory, protecting, saving, standing up, fighting for.

Victim suffer, harm, hurt, sicken, kill, injure, devastate, destroy, wipe out, flood, burn, choke, target, scapegoat, discriminate, marginalize, burden, price out, left behind, displaced, vulnerable, at risk, worst-hit, impacted, frontline, underserved, low-income, fenceline, victim, victims, screwed, shafted, ruined, wrecked, suffering, hurting, dying, starving, homeless, unfair, injustice, environmental justice, ej community, ej communities.

Examples of Tweets

In this subsection we provide some examples of tweets that were coded by the human coders, by the LLM mode, and our alternative NLP pipeline.

Table B.2: ***Illustrative Examples of Coder Agreement***

Tweet	Human 1	Human 2	LLM	NLP
"@Fissionary Correct, solar power doesn't create energy when no solar source, but peak power can be stored & recovered later=no fossil fuels."	Fossil industry (villain), green tech (hero)	Fossil industry (villain), green tech (hero)	Fossil industry (villain), green tech (hero)	Green tech (neutral)
"@Andrs_lk Basically. There are a million ways that corporations can be enticed to limit their environmental impact. Tax breaks for businesses producing less than x amount of tons of carbon annually, carbon tax, stronger environmental protections for lakes, forests, air, etc."	Corporations (villain), emission pricing (hero), regulations (hero)	Corporations (neutral), emission pricing (hero), fossil industry (villain), regulations (hero)	Corporations (neutral), carbon pricing (hero), regulations (hero)	Corporations (neutral), carbon pricing (neutral)
"@NBCPhiladelphia We MUST move away from fossil fuels and move into the future! Right here in PA, we can lead the nation in renewable energy!"	Fossil industry (villain), green tech (hero), us people (hero)	Fossil industry (villain), green tech (hero), us people (hero)	Corporations (neutral), fossil industry (villain), green tech (hero), us people (neutral)	Green tech (hero)

Notes: The table presents three illustrative tweets alongside their classifications by two human coders (Human 1, Human 2), the large language model (LLM), and the NLP dictionary-based classifier (NLP). Each entry reports the character(s) identified in the tweet and their assigned role in parentheses (villain, hero, victim, or neutral).

B.3 Frequency of Entities, Words, and N-Grams in the Tweets

We conclude this set of exercises by validating our character taxonomy against entities that emerge organically from the text, where named entities are noun phrases automatically extracted from text that refer to real-world objects such as individuals, organizations, and geopolitical actors. The aim is to *(i)* identify the most common entities in tweets classified as being about climate policy, and *(ii)* assign these entities to our ten characters. This allows us to assess coverage: if the entities that appear most frequently in the data map largely onto our pre-defined characters, this provides evidence that our taxonomy captures the actors and instruments that matter most in climate policy discourse. Given we are already aware that a subset of tweets contains no character from our list, we expect strong but not complete coverage.

We apply spaCy's pre-trained NER model to extract named entities from cleaned tweet text, retaining entities belonging to the following semantic categories: PERSON for named individuals such as politicians, activists, and CEOs; ORG for organizations, companies, agencies, and institutions; GPE for geopolitical entities including countries, states, and cities; LOC for non-GPE locations and landmarks; EVENT for named events, conferences, and disasters; NORP for nationalities, religious, or political groups; FAC for facilities, buildings, and infrastructure; PRODUCT for products and technologies; and LAW for laws, treaties, and legal instruments. Each entity is then assigned to one

of our ten characters using GPT-4o-mini, with explicit guidance for ambiguous cases. Table B.3 reports the results.

Panel (a) of Table B.3 reports the 25 most frequent entities in our sample of climate-policy tweets, along with the character to which each entity maps and its frequency of appearance. The top entities include geopolitical actors (China, US), organizations (GOP, UN), and institutions (EPA, Congress), among others. When an entity does not map to one of our ten characters, we assign it to a residual *Other* category with a descriptive label — for example, Iowa maps to *Other: US State*. Panel (b) provides complementary information: for each of our ten characters and each *Other* category, it shows how many of the top 100 most frequent entities map into that group.

The exercise provides a reassuring picture. Of the top 25 most frequent entities, 17 map into one of our characters, covering roughly 70% of the most frequent entities. Similarly, roughly 65% of the top 100 most frequent entities map into one of our ten characters. Taken together, these results suggest that our pre-defined characters capture the dominant actors in climate-policy discourse. The entities that fall outside our taxonomy are largely peripheral – US states, minor organizations – rather than central participants in the political debate over climate policy.

Table B.3: *Top Entities and LLM Character Assignment*

(a) Top Entities and Assigned Character Category

Rank	Entity	Character (LLM)	Frequency
1	china	Developing Economies	6937
2	us	US People	5907
3	gop	US Republicans	5298
4	paris	Other: High-Income Countries	4296
5	california	Other: US State	4169
6	republicans	US Republicans	4138
7	epa	Regulations	3787
8	congress	Democrats/Republicans	3710
9	america	US People	3650
10	american	US People	3171
11	americans	US People	3108
12	democrats	US Democrats	2676
13	india	Developing Economies	2271
14	iowa	Other: US State	2193
15	texas	Other: US State	2136
16	republican	US Republicans	1791
17	senate	Democrats/Republicans	1649
18	bernie	US Democrats	1631
19	un	Other: High-Income Countries	1576
20	florida	Other: US State	1549
21	dems	Democrats	1397
22	europe	Other: High-Income Countries	1252
23	russia	Other: High-Income Countries	1170
24	dem	US Democrats	1149
25	the united states	US People	1123

(b) Summary by Character Category

Character	No. of Entities	Frequency	Share (%)
Developing Economies	10	12,616	12.3%
US Democrats	19	12,911	12.5%
US Republicans	8	12,696	12.3%
Corporations	1	826	0.8%
US People	5	16,959	16.5%
Regulations	2	4,805	4.7%
Democrats/Republicans	3	5,951	5.8%
Other: High-Income Countries	14	12,347	12.0%
Other: US State	23	16,589	16.1%
Other	15	7,265	7.1%
Total	100	102,965	100.0%

Notes: Panel (a) reports the top 25 most frequently mentioned named entities in the tweets classified as being about climate change policy through our classification pipeline. The entities are then classified into character categories using a large language model (GPT-4o-mini). Entities were extracted from climate-related tweets posted in the US between 2010 and 2021 using spaCy's Named Entity Recognition (NER). Each entity was then classified by the LLM into one of the 10 character categories used in our narrative framework, with additional categories for: (i) *Other: High-Income Countries*, (ii) *Other: US State*, (iii) *Democrats/Republicans* for government institutions whose partisan affiliation is time-dependent, and (iv) *Other* for entities that do not fit any category. Frequency counts represent the total number of entity mentions across the analyzed sample of relevant climate tweets. Panel (b) summarizes the distribution of the top 100 entities across character categories.

We finalize this exercise with a simpler yet informative approach. We implement a weighted log odds analysis with an informative Dirichlet prior, which identifies words and phrases used significantly more often in tweets mentioning a particular character compared to all other tweets, while accounting for sampling variability through Bayesian shrinkage. For each character, we partition the corpus into a target set (tweets where that character is mentioned) and a reference set (all remaining tweets). The Dirichlet prior regularizes estimates for rare terms, preventing spurious associations. We rank terms by their z-scores, which capture both the magnitude and statistical reliability of each association. Table B.4 presents the top five most distinctive unigrams and bigrams for each human and instrument character. The resulting semantic profiles are intuitive and consistent with the roles assigned by our pipeline, providing further face validity for the taxonomy.

Table B.4: ***Ranking of Most Distinctive Words (Weighted by Their Frequency)***

(a) Human Characters

	Developing		Democrats		Republicans		Corporations		People	
	Word	Bigram	Word	Bigram	Word	Bigram	Word	Bigram	Word	Bigram
1	china	china india	obama	green deal	trump	stop denying	oil	fossil fuel	americans	human health
2	countries	per capita	president	national emergency	science	denying science	industry	oil gas	consumers	sustainable policy
3	india	waving hand	vote	showing americas	denying	national emergency	companies	oil industry	action	policy puts
4	poor	developed countries	deal	health care	president	congress act	solar	fuel industry	peace	health peace
5	nations	women girls	biden	declare national	administration	science congress	gas	ralph nader	kids	consumers human

(b) Instrument Characters

	Pricing		Regulations		Fossil		Green Tech		Nuclear	
	Word	Bigram	Word	Bigram	Word	Bigram	Word	Bigram	Word	Bigram
1	tax	gas tax	fracking	green deal	gas	fossil fuels	power	oil industry	nuclear	wind solar
2	taxes	action lead	regulations	sustainable policy	oil	fossil fuel	renewable	solar power	power	power plants
3	price	reform plan	green	policy puts	fuel	natural gas	solar	fossil fuels	plants	solar wind
4	credits	plan city	ban	peace dennis	industry	oil gas	wind	sustainable policy	free	natural gas
5	market	charge action	deal	dennis kucinich	fuels	oil companies	clean	consumers human	wind	nuclear war

Notes: The tables report the top 5 most mentioned terms within each character category, for all ten characters used in and for our classification. The terms are ranked by the weighted log-odds statistic with a Dirichlet prior. The ranking is based on the z-score, which indicates how distinctive a term is for the target character relative to all other tweets. Background probabilities use the full corpus. Terms with fewer than 5 total occurrences were omitted.

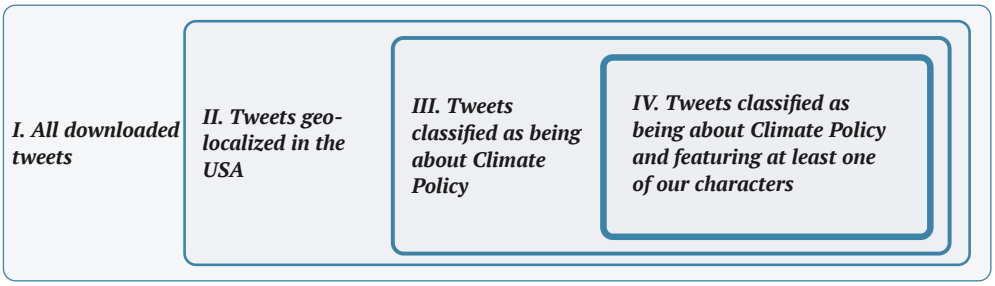
C Additional Output: Twitter Data

C.1 Descriptive Statistics on Tweets and Users

In this section of Appendix C, we provide an overview of the datasets used in the analysis, followed by descriptive statistics on the tweets and their authors. We begin with Figure C.1 that provides a visualization of the most relevant datasets. Dataset I contains all tweets downloaded through the Twitter API v2. Dataset II is a subset of I, restricted to tweets geolocated in the US. Dataset III further restricts to tweets classified by our pipeline as relevant to climate policy. Dataset IV adds the requirement that at least one of our ten characters is present.

We report descriptive statistics for datasets II, III, and IV – the three samples used in the analysis (in the paper and in the appendix). Table C.1, Table C.2, and Table C.3 show descriptive statistics respectively for datasets II., III., and IV. of Figure C.1. 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.

Figure C.1: ***Structure of the Dataset Used in the Analysis***



Notes: The figure shows the nested structure of datasets used in the analysis. Dataset I: all tweets downloaded via the Twitter API using climate-related keyword filters, covering every Saturday and one randomly selected day per month, 2010–2021. Dataset II: US-geolocated tweets (subset of I). Dataset III: tweets classified as relevant to climate policy (subset of II). Dataset IV: tweets featuring at least one of our ten characters (subset of III).

Table C.1: ***Features of All Tweets Localized in the USA (Dataset II., United States, 2010-2021)***

	Mean	Median	St. Dev.	Min.	Max.
Virality					
No. of Retweets (Virality)	2.2	0	220	0	194,217
No. of Likes	11	0	1,139	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	.8	.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	1,151,521				

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 4 is the reference section, where we describe and discuss the data used in the analysis.

Table C.2: ***Features of All Climate Policy-Relevant Tweets Localized in the USA (Dataset III., United States, 2010-2021)***

	Mean	Median	St. Dev.	Min.	Max.
Virality					
No. of Retweets (Virality)	3.6	0	152	0	29,526
No. of Likes	18	0	1,174	0	489,375
Tweet's Characteristics					
No. of Words	28	25	13	1	64
No. of Hashtags	.44	0	1	0	26
No. of Mentions	1.9	1	5.3	0	51
Quality of Text					
Lexical Density	.81	.82	.083	0	1
Type/Token Ratio	.93	.94	.062	0	1
Reading Ease	56	58	22	-1148	118.68
Educa. Needed to Comprehend Text	12	11	4.9	0	54.62
Emotions					
Avg. Count of Joy Words	.045	0	.079	0	1
Avg. Count of Surprise Words	.027	0	.067	0	1
Avg. Count of Trust Words	.11	0	.16	0	1
Avg. Count of Anger Words	.048	0	.085	0	1
Avg. Count of Disgust Words	.029	0	.069	0	1
Avg. Count of Fear Words	.16	.11	.22	0	1
Avg. Count of Sadness Words	.052	0	.093	0	1
No. of Observations	370,234				

Notes: The table reports descriptive statistics for the dataset of tweets about climate policy. 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 4 is the reference section, where we describe and discuss the data used in the analysis.

Table C.3: ***Features of Policy-Relevant Tweets Localized in the USA with at least One Character (Dataset IV., United States, 2010-2021)***

	Mean	Median	St. Dev.	Min.	Max.
Virality					
No. of Retweets (Virality)	3.8	0	160	0	29,526
No. of Likes	19	0	1,269	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	.08	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	.16	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	310,230				

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 4 is the reference section, where we describe and discuss the data used in the analysis.

In Table C.4, we show descriptive statistics about users that posted the tweets used in our analysis, corresponding to dataset IV., as described above. 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 can be located only at the national level, they are 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: ***Features of Users That Posted Tweets Containing at Least One Character (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	8,848	432	424,423	0	133,245,480
No. of Following	1,701	667	6,409	0	850,382
No. of Tweets	25,090	8,037	57,924	1	3,671,805
Profile Description					
Share of Democrats	.14	0	.35	0	1
Share of Republicans	.033	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	152,482				

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 4 is the reference section, where we describe and discuss the data used in the analysis.

C.2 Additional Descriptive Output on Narratives' Frequency and Virality

This section provides additional descriptive output on our pipeline's tweet classifications, specifically, (i) how GPT classifies tweets into narrative categories and (ii) how the virality outcomes used in the analysis are distributed.

We begin with Figure C.2, which summarizes tweet classifications as shares. Panel A shows the higher-level classification into four categories: *Irrelevant*, *Assert Issue Climate Change*, *Deny Climate Change*, and *Climate Change Policy*. These correspond, respectively, to tweets irrelevant to the discussion, tweets that assert the importance of the issue without engaging with policy or characters, tweets that deny climate change, and tweets relevant to climate change policy. In terms of Figure C.1, Panel A covers Dataset II, with the *Climate Change Policy* subset corresponding to Dataset III.

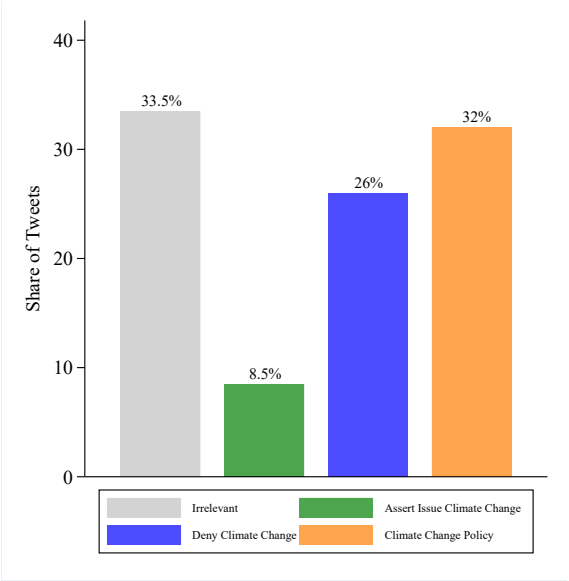
Panel B of Figure C.2 breaks down Dataset III by the number of character-roles per tweet. A tweet may contain none of our ten characters, characters in neutral roles only, or characters cast as hero, villain, or victim. The grey bar shows tweets about climate policy that feature none of our characters — we exclude these from virtually all analyses. Our analysis thus uses the remaining tweets, those featuring at least one character whether in a neutral or drama triangle role, corresponding to Dataset IV.

Figure C.2a shows that among US tweets, roughly 33% are classified as irrelevant — too unclear, too short, or unrelated to climate change despite containing keywords from our filter. Tweets classified as discussing climate change policy account for about 32%. Much of the remaining conversation consists of tweets that simply assert or deny man-made climate change (green and blue bars). These tweets do not engage with policy; instead, they reflect fixed and polarized positions in the broader debate.

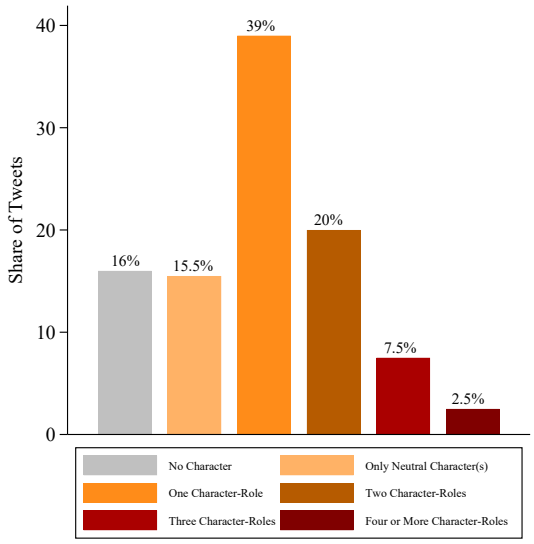
We prompt GPT to identify characters only for tweets in the *Climate Change Policy* category. Among those, tweets featuring at least one character form the set of `relevant` tweets (310,230 tweets). Figure C.2b breaks down this set. In 18.5% of tweets, characters appear in neutral form only — not assigned any drama triangle role. This subset serves as the comparison group in most of our analysis and underpins our estimates of the effect of political narratives. The remaining tweets contain at least one character-role: 46% include one, 23% two, 9% three, and 3% four or more.

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.

Below, we provide additional details on the distribution of our character-roles, exploring how characters are classified across roles and neutral framing. Table C.5 displays the distribution of character-roles in shares, excluding character-roles with fewer than 100 occurrences in the full set of classified tweets. Table C.6 provides the distribution in absolute counts, including all character-roles, also those below 100. In both tables, the fifth column represents the overall share or number of times a character appears in Dataset IV (as in Figure C.1). This is broken down to show how many times that character appears as hero, villain, victim, or in a neutral role.

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 (87). 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 occurrences happen 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 5,322 and 18,180 occurrences, respectively.

Table C.5: ***Share of Character-Roles in Relevant Tweets (United States, 2010 - 2021)***

(a) Human Characters

	Hero	Villain	Victim	Neutral	Total
Developing Economies	0.13	1.20	0.91	0.20	2.44
US Democrats	5.71	1.82	0.06	1.36	8.94
US Republicans	0.12	9.49	.	1.58	11.19
Corporations	0.99	8.14	0.06	7.84	17.03
US People	4.00	0.55	3.09	9.64	17.28

(b) Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emission Pricing	2.03	1.24	.	1.19	4.47
Regulations	3.47	1.36	.	3.17	8.01
Fossil Industry	0.06	11.76	0.04	1.60	13.46
Green Tech	11.52	0.83	.	3.18	15.54
Nuclear Tech	0.86	0.32	0.02	0.44	1.64

Notes: The table summarises the distribution of character-roles across `relevant` tweets (those with $\sum_k m_{ik} \geq 1$), expressed as shares. Only character-roles appearing at least 100 times over 2010–2021 are shown; excluded cells are marked with dots. Each row decomposes a character's overall share in the sample by role. "Neutral" denotes characters for whom $m_{ik} = 1$ and $\sum_r r_{ikr} = 0$; multiple character-roles may co-occur within a tweet. Absolute counts, including excluded categories, are reported in Appendix Table C.6.

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

(a) Human Characters

	Hero	Villain	Victim	Neutral	Total
Developing Economies	780	7,044	5,322	1,176	14,322
US Democrats	33,538	10,706	341	7,964	52,549
US Republicans	686	55,774	95	9,293	65,848
Corporations	5,791	47,868	374	46,069	100,102
US People	23,510	3,226	18,180	56,666	101,582

(b) Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emission Pricing	11,957	7,315	11	6,976	26,259
Regulations	20,422	8,009	29	18,660	47,120
Fossil Industry	362	69,098	236	9,401	79,097
Green Tech	67,715	4,903	87	18,700	91,405
Nuclear Tech	5,082	1,884	107	2,587	9,660

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.

We proceed with Table C.7. This figure complements the paper Figure 4, which shows the variation in the occurrence of each single character-role of our analysis in the period 2010-2021. The table provided below displays the average yearly 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	23.04	9.82	10.41	42.18
Fossil Industry-Villain	16.91	2.72	10.95	20.61
U.S. Republicans-Villain	11.82	5.85	2.77	23.70
Corporations-Villain	10.65	2.11	6.85	14.96
U.S. Democrats-Hero	6.69	2.80	3.21	12.79
U.S. People-Hero	5.21	1.16	3.59	7.98
Regulations-Hero	4.55	1.10	2.79	6.25
U.S. People-Victim	3.41	1.57	1.29	6.11
Emiss. Pricing-Hero	3.23	0.79	2.09	4.63
Emiss. Pricing-Villain	2.15	0.74	1.21	3.71
Corporations-Hero	2.07	1.88	0.89	7.82
U.S. Democrats-Villain	1.94	0.97	0.79	3.98
Regulations-Villain	1.76	0.37	1.28	2.39
Developing Economies-Villain	1.43	0.46	0.75	2.26
Developing Economies-Victim	1.20	0.27	0.83	1.77
Green Tech-Villain	1.19	0.38	0.66	1.84
Nuclear Tech-Hero	0.93	0.42	0.30	1.66
U.S. People-Villain	0.61	0.26	0.23	1.05
Nuclear Tech-Villain	0.56	0.37	0.34	1.69
Developing Economies-Hero	0.22	0.08	0.09	0.39
U.S. Republicans-Hero	0.14	0.08	0.01	0.31
Corporations-Victim	0.10	0.07	0.00	0.27
U.S. Democrats-Victim	0.06	0.04	0.01	0.14
Fossil Industry-Hero	0.06	0.04	0.01	0.13
Fossil Industry-Victim	0.05	0.02	0.02	0.09
Nuclear Tech-Victim	0.02	0.02	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, for the period 2010-2021. 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. Subsection 5.3 is the reference section, where we describe and discuss the evolution of character-role combinations over time. The reference plot is Figure 4.

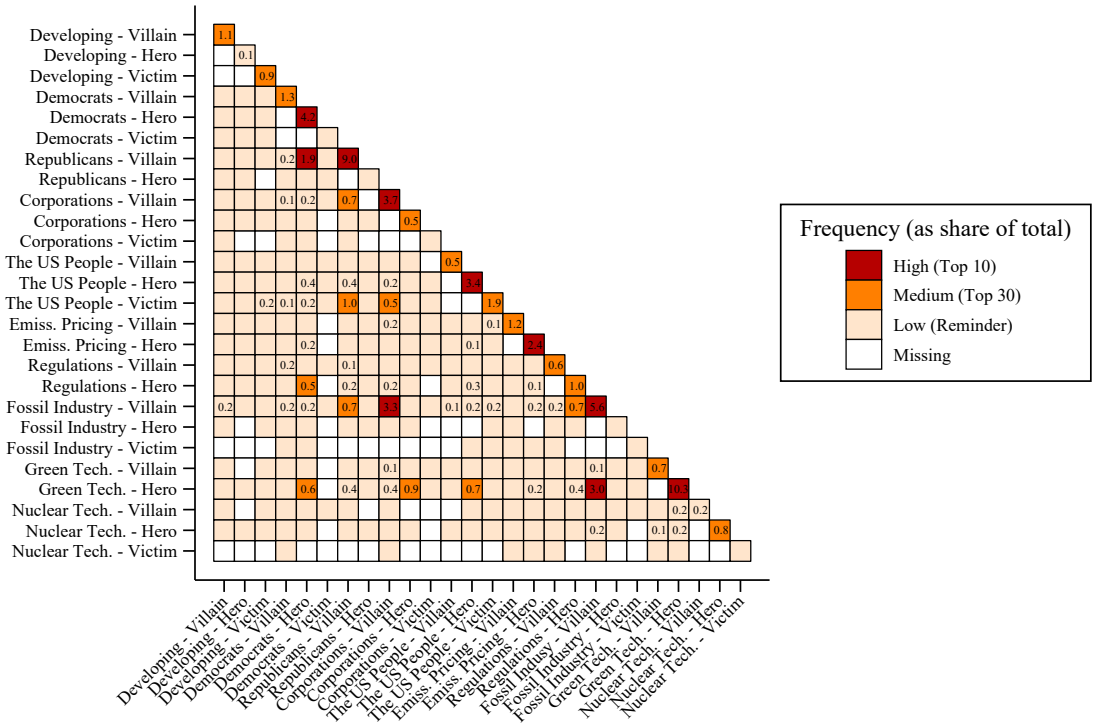
Below, we complement the information reported in the paper Figure 5. Figure C.3 shows the same plot, but complemented by annotating the most frequent single character-roles – in the diagonal – or configurations of two character-roles – off the diagonal. Additionally, we provide an in-depth explanation on how the measures in the matrix are computed.

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 $N_i = \sum_{k \in K} \sum_{r \in \mathcal{R}} r_{ikr}$. 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.

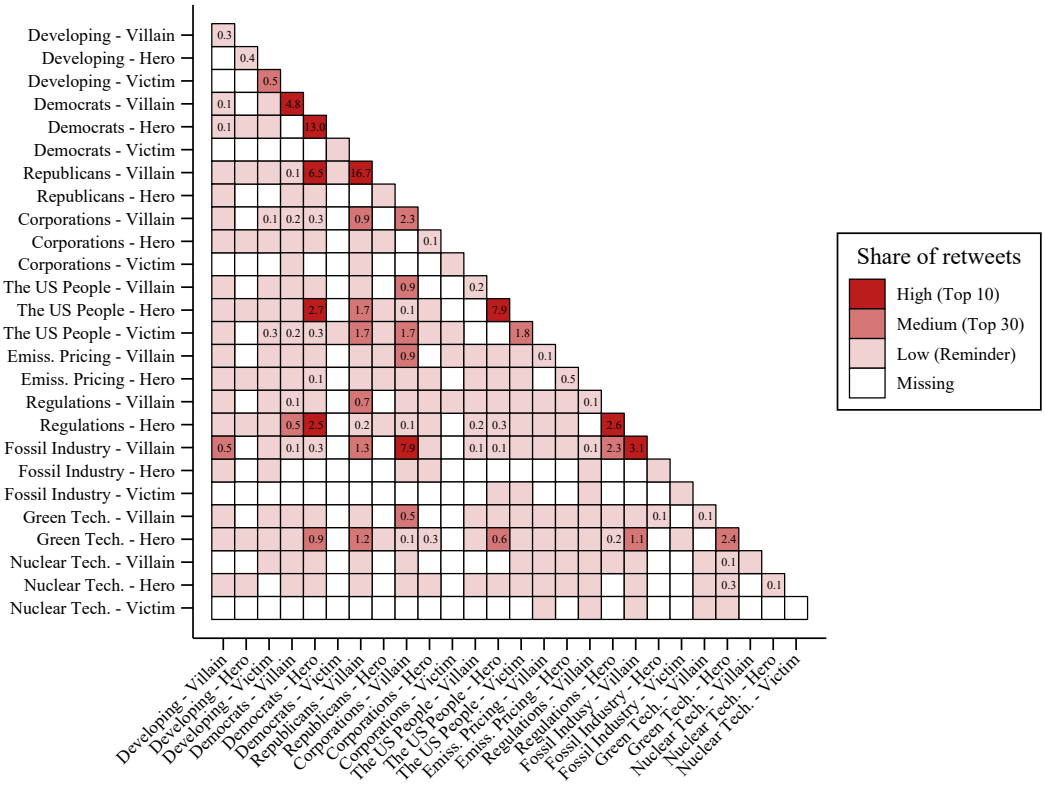
For what concerns Figure C.4, we follow the same logic but replace tweet counts with retweet counts. That is, we compute a weighted matrix $m^{(w)}_{ij}$ where each entry reflects the share of total retweets received by tweets in $\mathcal{D}^{\text{rel}}_{[1,2]}$ that contain a given character-role 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 5 is the reference plot.

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

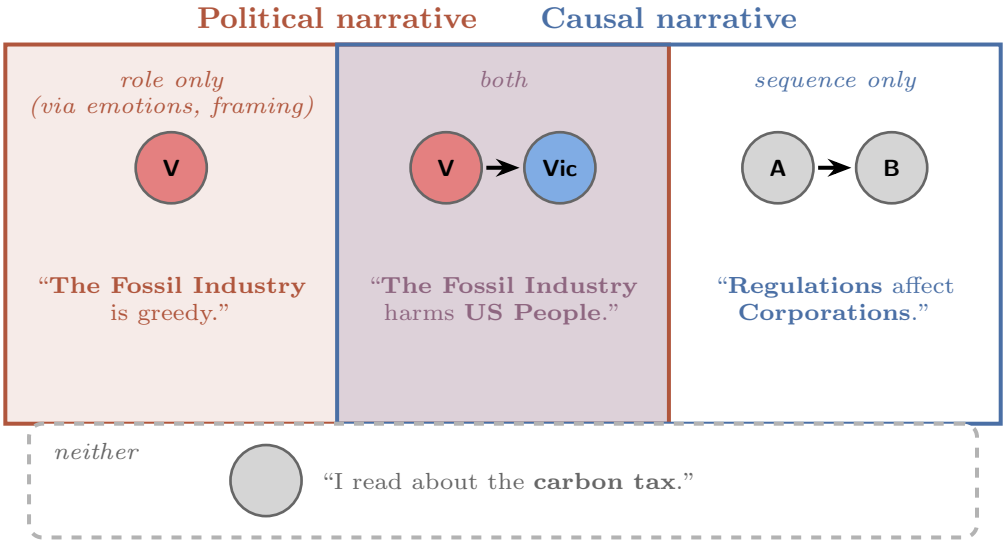


Notes: The figure shows the retweet share of each character-role appearing either alone or in combination with another role, among relevant tweets containing one or two roles. Retweet rates are computed as the share of total retweets received by a given role (or pair) relative to all retweets of tweets 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%).

C.3 Political and Causal Narratives

Our framework relates to, but does not coincide with, the causal-narrative approach. A text is a political narrative in our sense when at least one character is cast in a drama-triangle role, and a causal narrative in the sense of Eliaz and Spiegler (2020) when it links characters through cause and effect. These two criteria are logically independent and partition texts into four regions (Figure C.5): role assignment without a causal chain, causal structure without roles, both, or neither. The role-only region dominates real political speech, which is why we anchor the definition on roles while remaining compatible with causal-graph representations layered on the same characters.

Figure C.5: ***Political and Causal Narratives: Four Regions of the Definitional Space***



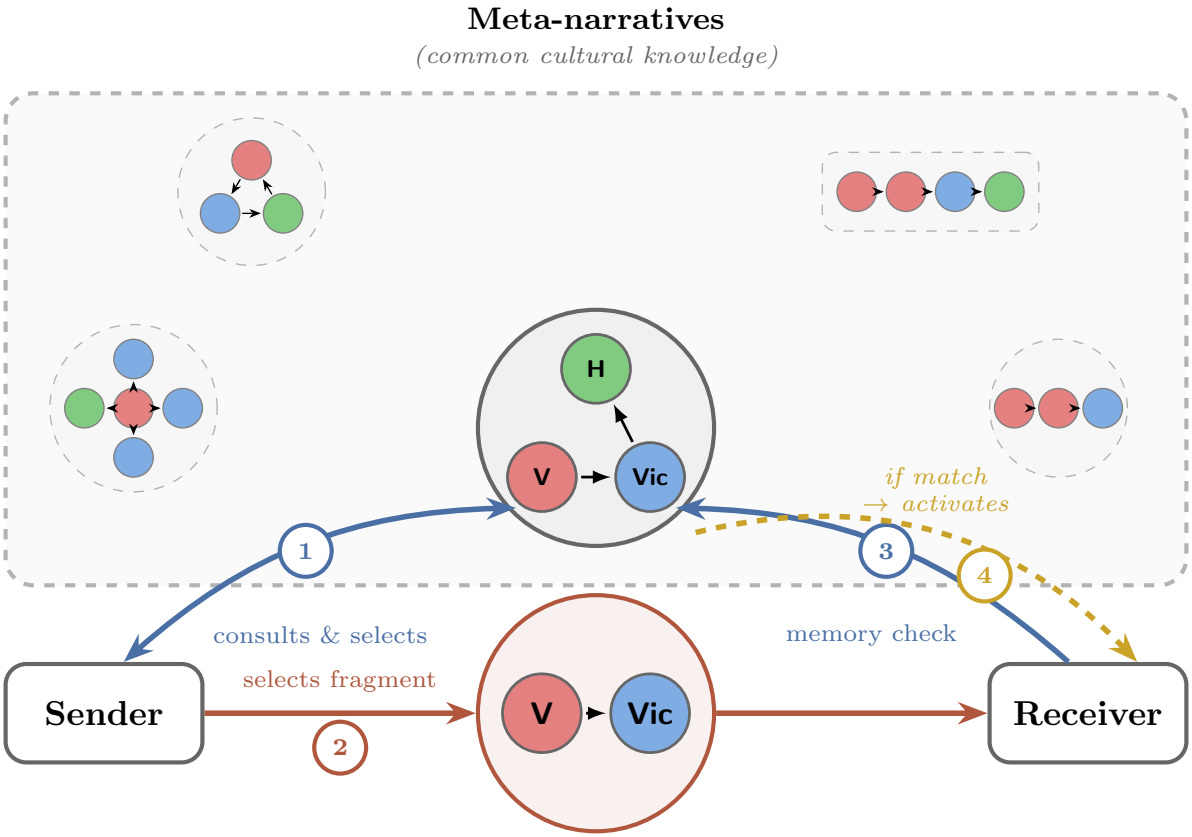
Notes: The figure visualizes the four regions implied by combining role assignment $r(\cdot)$ and causal structure E. A text is a *political narrative* (this paper) iff at least one character is assigned a drama-triangle role, $r(k) \in \{\text{hero}, \text{villain}, \text{victim}\}$ (left rectangle). It is a *causal narrative* (Eliaz and Spiegler (2020) and follow-ups) iff $E \neq \emptyset$ (right rectangle). The intersection (*both*) contains texts that simultaneously assign roles and articulate causal links between characters. *Role only* captures fragments that assign roles via emotions or framing without a causal chain — empirically the most frequent case in our sample. *Sequence only* captures statements that link characters causally without assigning any role (e.g., factual or counterfactual claims). *Neither* contains texts that mention characters only neutrally and without sequential structure. The two definitions overlap substantially but neither nests the other.

C.3.1 Meta-narratives and Fragments

Role-only fragments are not impoverished narratives but compressed ones. A short message casting a single character in a role typically gestures at a larger, shared meta-narrative the audience already holds (Figure C.6). The sender selects a fragment of a familiar story; the receiver, recognizing the character and its role, reconstructs the fuller narrative from memory, including causal links the fragment never states. This is why coding characters and roles recovers the essence of a narrative even when the text is sequence-free: the fragment is a retrieval cue for a meta-narrative that may

or may not itself be causal.

Figure C.6: ***Meta-narratives, Fragments, and Receiver Activation***



Notes: The figure illustrates how role-only fragments transmit narrative content by activating shared cultural knowledge in the receiver. The grey dashed box (top) is the pool of *meta-narratives* the receiver already holds; the focal meta-narrative (centered) is a drama-triangle structure (hero, villain, victim), and the surrounding smaller meta-narratives illustrate that meta-narratives vary in shape and node count. The sender consults the pool and selects a focal meta-narrative (arrow 1), then transmits a *fragment* of it to the receiver (arrows 2): here the fragment is a $V \rightarrow Vic$ pair, omitting the other characters and the full causal chain. The receiver performs a memory check against their own repertoire (arrow 3) and, if a match is found, the full meta-narrative is activated (arrow 4, dashed). This mechanism explains why role-only fragments convey rich narrative content without articulating a full causal chain: the fragment hooks into a meta-narrative the receiver already holds. The meta-narratives can be causal — as the figure shows with arrows inside each — but our framework does not require them to be; fragments can hook into either causal or non-causal shared narratives.

C.3.2 Causal Connections in Political Narratives

Our narrative framework does not require characters to be causally connected, and our main estimates do not condition on the presence of such a link. A natural question is whether causal structure between characters is nevertheless identifiable from tweet text, and how often it actually appears in our sample. The DAGs in Figure C.7 provide a representation of how the causal links

integrate within the political narrative framework.

Figure C.7: Causal Extension of the Political-Narrative Classification

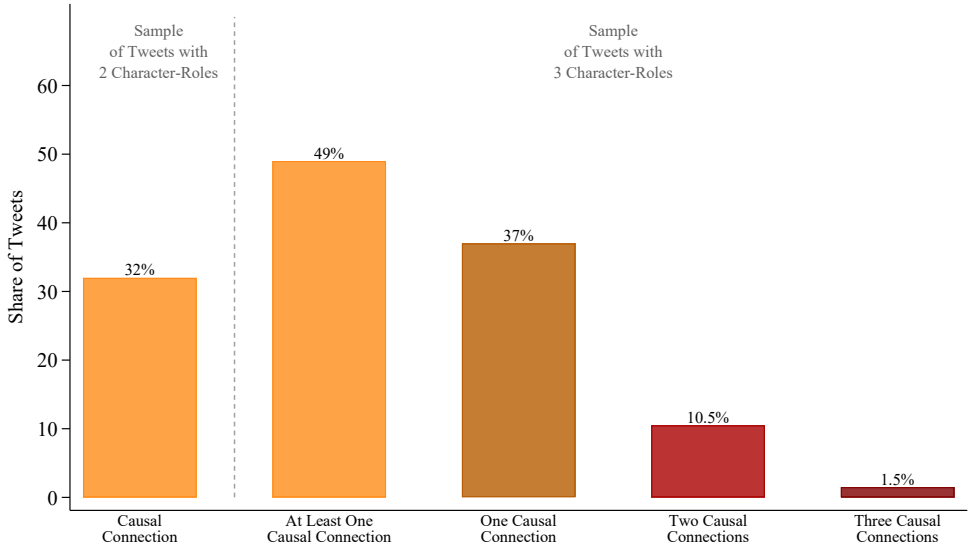
Notes Stages 1–2 identify the characters and their roles (hero **H**, villain **V**, victim **Vic**); the third stage adds, for each pair, a directed edge labelled with an active verb when the tweet describes a causal connection. Panels (a)–(b) are dyadic cases: the first with a clear verb in the sentence, the second with an inferred action. Panel (c) is a triadic example where only one of three pairs yields an edge; Trump's paper-towel response is set against Musk's offer but does not act on either Musk or solar power. Panel (d) shows a chain: Progressives write a climate bill that targets fossil fuels; the Democrats–Fossil Fuels pair is empty because the action runs through the policy, not directly.

To address this, we use the same LLM model to ask, for every pair of characters returned by our stage-2 classification, whether the tweet itself describes a causal connection in which one character actively acts on the other. The prompt is deliberately conservative. It requires that the

relation be summarised by a short active verb whose meaning follows from what the tweet actually says, rules out comparisons, juxtapositions, and third-party narration in which the speaker rather than the characters does the acting, and asks the model to abstain whenever the link can only be sustained by background knowledge. The roles assigned in stage 2 are not shown to the model, so any direction it returns has to be recovered from the text alone. We apply the procedure to 500 tweets with exactly two character-roles and 500 tweets with exactly three character-roles, drawn at random from the relevant tweets in our main sample.

Figure C.8 reports the results. The first bar refers to the two-character subsample and gives the share of tweets in which the model identifies a causal connection between the two characters. The remaining bars refer to the three-character subsample: the share of tweets with any causal connection across the three possible pairs, and then the share with exactly one, exactly two, and exactly three directed connections.

Figure C.8: ***Measuring Causal Connections on top of Political Narratives***



Notes: The figure shows the share of tweets in which our LLM-based classifier identifies a causal connection between characters. Two subsamples are used, separated by the dotted vertical grey line. The leftmost bar refers to a random sample of 500 relevant tweets containing exactly two character-roles, and reports the share of tweets in which the model identifies a causal connection between the two characters. The remaining four bars refer to a random sample of 500 relevant tweets containing exactly three character-roles. Classification is performed with `gpt-4o-mini`. For each pair of characters returned by our stage-2 classification, the model is asked whether the tweet describes a causal relation in which one character actively acts on the other. The roles assigned in stage 2 are not shown to the model, so any direction the model returns must be recovered from the text alone.

Causal connections are present in roughly one third of two-character tweets and in close to half of three-character tweets, so explicit causal structure between flagged characters is far from universal in our dataset. Within the three-character subsample, partial graphs clearly dominate. Tweets in which the model identifies exactly one causal connection are the most common case, tweets with two connections appear in only about one in ten cases, and tweets in which all three

pairs are causally connected account for a negligible share of the sample. We read this as evidence that political narratives on Twitter typically rely on characters and roles without spelling out a full causal mechanism, which is consistent with our choice to focus on the political narrative dimension in the main analysis rather than condition on the presence of an explicit causal link.

D Regression Tables Underlying the Main Figures: Twitter Data

In this section, we complement the main results presented in Section 6. For each graphical exhibit in the paper section, we report tables displaying additional information on the underlying models. Table D.1 lists the coefficients displayed in Figure 7a. Table D.2 lists the coefficients displayed in Figure 7b. Table D.3 reports the coefficient in Figure 8. Table D.4 shows the marginal effects plotted in Figure 9b.

Table D.1: ***Regression Results – Impact of Political Narratives on Virality***

Dependent Variable	**Retweets Count**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Political Narrative	0.660	0.725	0.598	0.481
	(0.252)	(0.249)	(0.131)	(0.137)
	[0.009]	[0.004]	[0.000]	[0.000]
Hour FE		✓	✓	✓
Year-Week FE		✓	✓	✓
Year-State FE		✓	✓	✓
Author Characteristics			✓	✓
Character FE				✓
Percentage Change	93.5%	106.5%	81.8%	61.8%
Mean Outcome (Control Group)	2.17	2.18	2.18	2.18
Pseudo R^2	0.004	0.257	0.622	0.635
Observations	310,230	310,016	310,016	310,016

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, year × calendar-week, 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. Plot of reference is Figure 7a.

Table D.2: ***Regression Results – Impact of Political Narratives on Virality***

Dependent Variable	**Retweets Count**								
	Coeff./SE/p-value								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Political Narrative	0.481	0.448	0.472	0.273	0.480	0.361	0.511	0.485	0.492
	(0.137)	(0.139)	(0.134)	(0.117)	(0.139)	(0.106)	(0.154)	(0.115)	(0.188)
	[0.000]	[0.001]	[0.000]	[0.019]	[0.001]	[0.001]	[0.001]	[0.000]	[0.009]
Hour, Year-Week, Year-State FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Author Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓
Character FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Control for Channel: Language Metrics		✓							
Control for Channel: Emotions/Valence			✓						
Sample: Full Climate Policy				✓					
Sample: Excluding Potential Bots					✓				
Algorithm: Before Change						✓			
Algorithm: After Change							✓		
Reach: Followers Below Average								✓	
Reach: Followers Above Average									✓
Mean Outcome (Control Group)	3.836	3.836	3.836	2.216	4.112	1.029	4.514	0.559	46.762
Pseudo R^2	0.635	0.676	0.638	0.566	0.634	0.696	0.644	0.382	0.682
Observations	310,016	310,016	310,016	1,151,521	287,359	60,312	249,675	287,604	21,978

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. Each column corresponds to one specification shown in Figure 7b. All specifications include hour, year × calendar-week, and year × state fixed effects, author characteristics, and character fixed effects. Standard errors are reported in parentheses and *p*-values in brackets.

Table D.3: ***Regression Results-Heterogeneity in the Impact of Character-Role Combinations on Virality***

Dependent Variable	**Retweets Count**		
	Sample: Hero	**Sample: Villain**	**Sample: Victim**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Hero - Villain	0.405	0.170	
	(0.194)	(0.208)	
	[0.037]	[0.414]	
Hero - Victim	-0.890		0.040
	(0.260)		(0.433)
	[0.001]		[0.926]
Villain - Victim		-0.093	0.560
		(0.276)	(0.194)
		[0.736]	[0.004]
Hero - Villain -	-0.676	-0.791	-0.017
Victim	(0.355)	(0.406)	(0.285)
	[0.057]	[0.051]	[0.952]
Multiple Heroes	0.029		
	(0.216)		
	[0.895]		
Multiple Villains		0.899	
		(0.269)	
		[0.001]	
Multiple Victims			0.391
			(0.449)
			[0.385]
Mean Outcome (Control Group)	3.438	3.374	2.447
Pseudo R^2	0.75	0.66	0.75
Observations	136,283	159,776	22,608

Notes: The table reports 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. Each column corresponds to a separate regression model. Column (1) 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. Column (2) follows the same structure for villains, with the reference group being tweets that feature a single villain character, while Column (3) does so for victims, using tweets with a single victim character as the baseline. All regressions include hour, year × calendar-week, and year × state fixed effects. Standard errors are clustered at the week level, covering all weeks in the study period. Plot of reference is Figure 8.

Table D.4: ***Marginal Effect of Adding Character-Roles***

Number of Character-Roles	**Predicted Retweets**	**p-value**
	Predicted	Pvalue
1	3.390955	4.80e-35
2	5.363294	5.41e-16
3	6.048259	3.33e-07
4	2.789567	.1353032
5	1.67027	.0237459
6	2.725572	.039429
7	1.958265	.3001824
8	.0000641	.3929085
Pseudo R^2	0.594	
Observations	310,230	

Notes: This table reports the marginal change in expected retweets associated with adding one additional character-role, showing how increasing narrative complexity affects virality. Because of technical reasons with the Poisson model, this regression excludes the year × state FEs. Reference plot is Figure 9b.

E Robustness Checks: Twitter Data

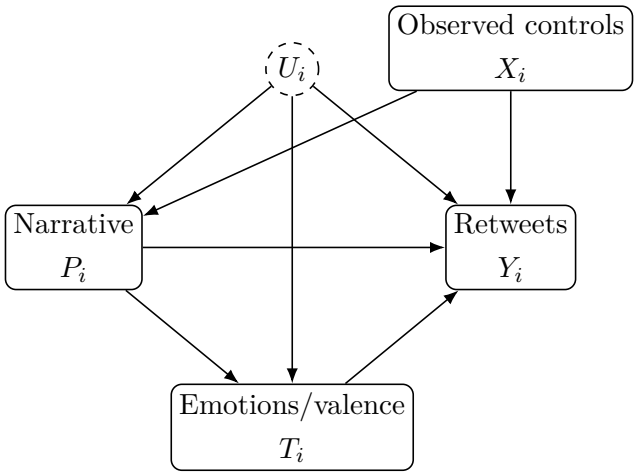
E.1 Visual Depiction of Language Metrics and Emotions/Valence Controls

Consider the visual scheme 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 $P_i \leftsquigarrow U_i \to Y_i$ 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 E.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.

E.2 Heterogeneity by Number of Followers

In this section, we ask whether users' reach on the platform, measured by their number of followers, shapes (i) the narratives they post and (ii) the virality of those narratives. A key limitation is that follower count is observed at the time of API extraction, not at the time the tweet was posted. The measure is therefore *ex post*. Still, it is the most informative proxy we have for profile reach.

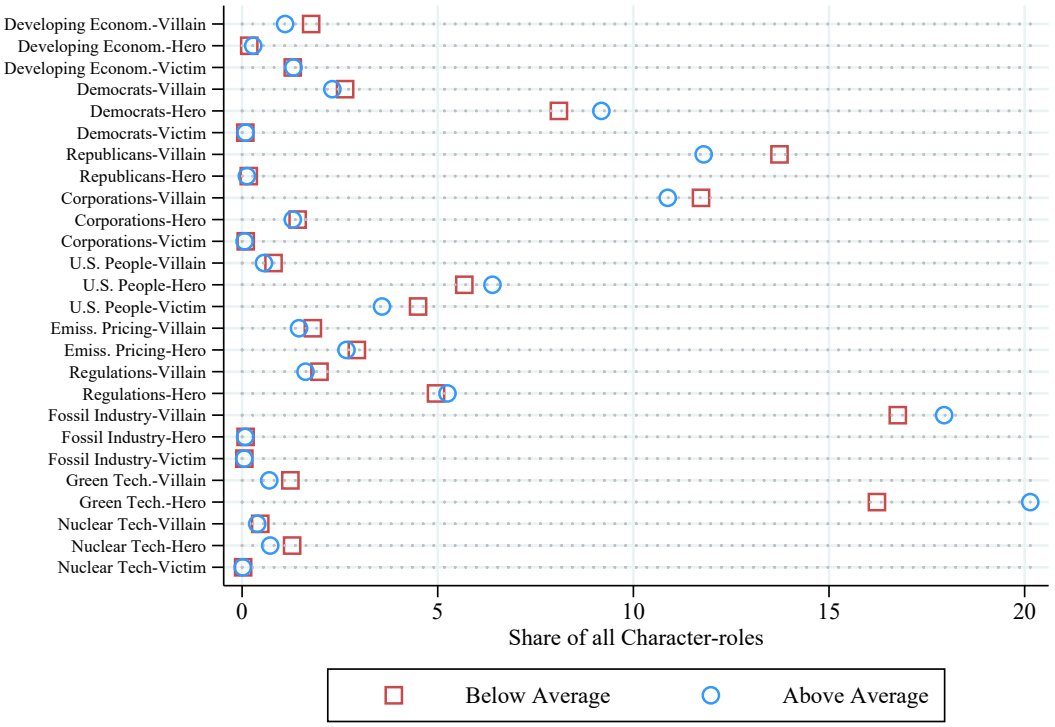
We split users into two groups: below- and above-average follower count (mean: 15,826). We first examine whether narrative content differs across reach levels, to ensure that any differences in virality do not mechanically reflect different narrative use. Figure E.2 plots, for both reach groups, the share of each character-role among all character-roles. Squares denote below-average reach profiles; circles denote above-average. We restrict attention to character-roles that appear at least 100 times in the full dataset.

Figure E.2 shows the share of each character-role in total character-role usage, separately for below- and above-average reach users. The distributions are broadly similar across the two groups. Above-average reach users post somewhat more Democrats-Hero and Fossil Industry-Villain narratives, though the differences are only 2–3 percentage points. The clearest exception is Green Tech-Hero, where above-average reach users account for a roughly 4 percentage point higher share. In the opposite direction, below-average reach users post a noticeably higher share of Republicans-Villain narratives, by about 2 percentage points. Overall, narrative composition appears broadly stable across reach groups, suggesting that differences in virality by reach are unlikely to be driven by differences in the narratives users choose to post.

Figure E.2 thus reassures us that there are no structural differences in narrative composition across users with different reach, measured in followers. However, similar narrative composition does not imply similar returns to narratives. We therefore turn to the second question raised above: do users with different reach obtain a different virality premium from posting narratives? We address

this in Figure 7b, which shows that both below- and above-average reach users exhibit a positive association between narrative use and tweet virality.

Figure E.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 below average followers count, circles represent those with above average followers count. The average number of followers is 15,826). 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.

E.3 Impact of Major Twitter Algorithm Changes

One crucial mechanism shaping virality of content on social media platforms is the algorithm that structures 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.

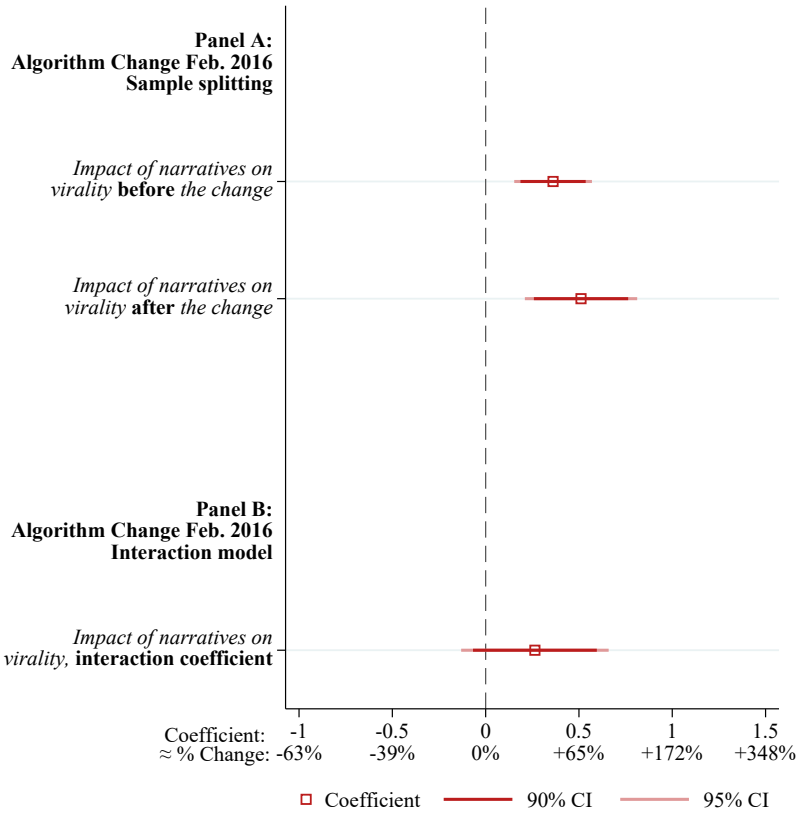
As BuzzFeed reported at the time, the algorithm was designed as "a way for Twitter to elevate popular content, and could solve some of Twitter's signal-to-noise problems. [...] The timeline will reorder tweets based on what Twitter's algorithm thinks people most want to see, a departure from the current 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 E.3.

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. This is the same we show also in the paper Figure 7b. 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.

Figure E.3: ***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, year × calendar-week, and year × state fixed effects. Standard errors are clustered at the week level.

E.4 Output Excluding Potential Bots

In this section of Appendix E, 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.

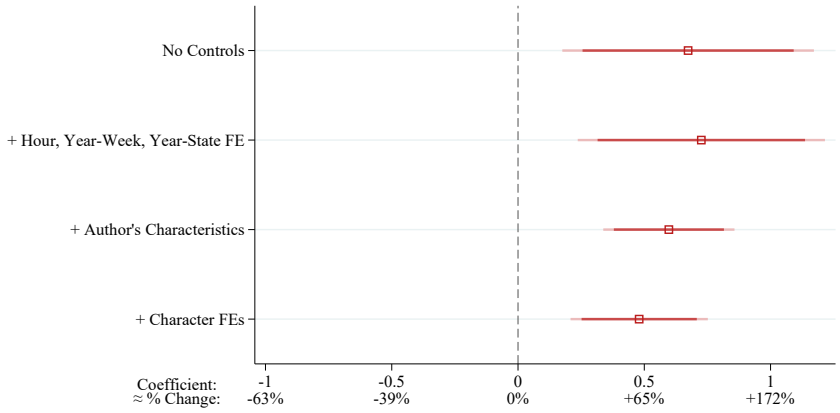
We implement two complementary strategies to detect potential bots based on both tweet content and user characteristics. Specifically, we define tweets as originating from potential bots if they meet one or both of the following two conditions:

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, most likely indicating that we capture different kinds of bots successfully. Among our relevant tweets, a total of 16,851 tweets were identified through definition 1., a total of 5,923 through definition 2., while only 112 overlap. We exclude these tweets in Figure E.4 which reproduces the paper Figure 7. Figure E.4 plots the coefficients from the same models used in the paper, but excludes tweets potentially posted by bots. 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.

Figure E.4: ***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 E.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, year × calendar-week, 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.

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 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.

E.5 Alternative Standard Errors Clustering

In the paper Figure 7b we provide evidence that clustering standard errors in different ways does not change substantially the main conclusions of the paper. 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.

E.6 Alternative Estimation Methods

Our main outcomes (retweets in the paper, and likes and replies in the appendix) are highly skewed count variables: many observations are zero or very small, while a small number are extremely large. This feature is typical of social-media engagement and complicates standard linear modeling choices. A common approach in applied work is to log-transform the outcome and estimate with OLS. This can work when outcomes are strictly positive, but our data contain many zeros, making log transformations ad hoc.

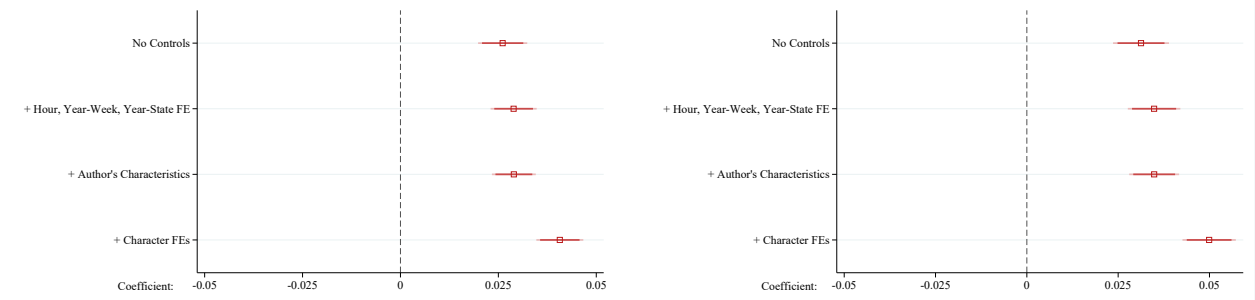
Recent work by Chen and J. Roth (2024) shows that common fixes such as $\log(y+1)$ can introduce non-trivial distortions. Adding a constant to accommodate zeros implicitly rescales the outcome, so estimated effects become a function of an arbitrary scaling factor (denoted "a" in Chen and J. Roth (2024)) determined by the chosen constant and by the distribution of y. In the extreme cases, treatment effects can be inflated or decreased simply by changing this constant. The broader concern is familiar from settings where units are arbitrary (e.g., hours worked): rescaling the dependent variable should not change substantive conclusions, yet with log-plus-constant transformations it can. Although this issue is less stark for count outcomes such as retweets or likes, the same instability remains.

For these reasons, we use Poisson Pseudo-Maximum Likelihood (PPML) as our preferred estimator. PPML is well suited to count outcomes, accommodates zeros without transformation, and delivers coefficients that can be interpreted like semi-elasticities, preserving a direct link between estimated effects and percentage changes in engagement. To facilitate comparison and to clarify what drives our main results, we also report some alternative specifications: (i) OLS after applying $\log(y+1)$ transformation, (ii) OLS after applying $\text{asinh}(y)$ transformation, and (iii) the piecewise approach proposed by Chen and J. Roth (2024), which separates extensive- and intensive-margin effects.

Linear Models in Levels and with Transformations We first replicate the analysis using OLS following the common transformation-based approach and estimate OLS models using $\log(y+1)$ and $\text{asinh}(y)$. Figure E.5a and Figure E.5b reproduce the main results from Figure 7. The qualitative conclusions match the PPML results: coefficients are positive and statistically significant at least at the 1% level. Overall, sign and significance are highly robust.

Figure E.5: ***Impact of Political Narratives on Virality-Log and Asinh Transformation***

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



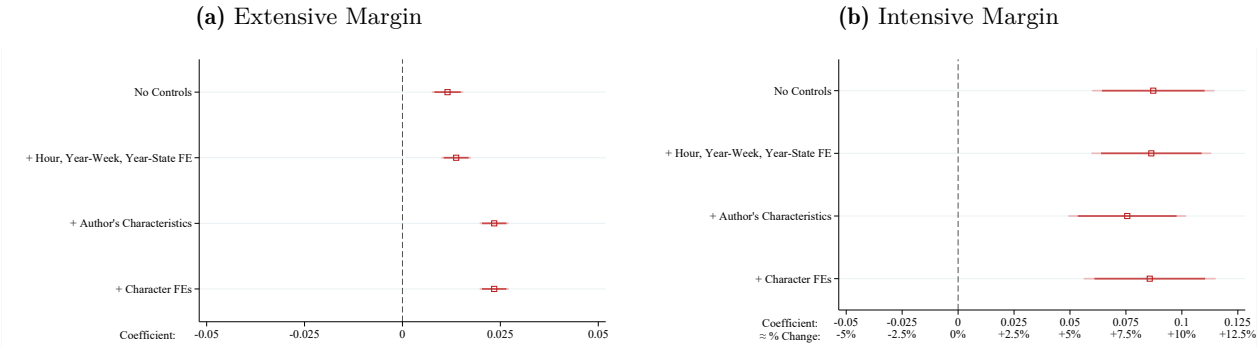
Notes: The figure shows the coefficients of OLS regressions testing the effect of featuring at least one character-role, compared to featuring characters only in a neutral role, on virality, measured as the count of retweets. In Figure E.5a the dependent variable is log transformed after adding one unit. In Figure E.5b 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.

Extensive and Intensive Margin Model Following Chen and J. Roth (2024), we also implement a piecewise approach that separates the probability of any engagement (extensive margin) from engagement conditional on being positive (intensive margin). We define two transformed outcomes:

$$y^{*1} = \begin{cases} y = 1 & \text{if } y > 0 \\ y = 0 & \text{if } y = 0 \end{cases} \qquad y^{*2} = \begin{cases} log(y) & \text{if } y > 0 \\ . & \text{if } y = 0 \end{cases}$$

The extensive-margin model uses y^{*1} and includes all observations, capturing whether narratives increase the likelihood of any engagement. The intensive-margin model uses y^{*2}, restricts to $y > 0$, and estimates OLS on $\log(y)$; coefficients can be interpreted as semi-elasticities, capturing changes in engagement intensity conditional on receiving some engagement. Figure E.6a and Figure E.6b report the extensive and intensive estimates, respectively. Political narratives increase both the probability of any retweet and the intensity of retweets conditional on being positive.

Figure E.6: ***Impact of Political Narratives on Virality-Extensive and Intensive Margins***



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 E.6a the dependent variable is first transformed to have $Y = 1$ if $Y > 0$, then a Linear Probability Model is applied. In Figure E.6b 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 E.6b 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 $\times$ 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.

Summary PPML remains our preferred specification given the distribution of the outcomes and the presence of many zeros. The alternative models nevertheless confirm that the core finding – a positive virality premium of political narratives – is robust to a range of common estimation strategies. At the same time, the comparison illustrates why transformation-based OLS approaches can be sensitive in zero-heavy settings, reinforcing the case for estimators tailored to count data.

E.7 Impact of Political Narratives on Popularity

In this section of Appendix E, we check the robustness of our results by examining how narratives affect tweets' 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.

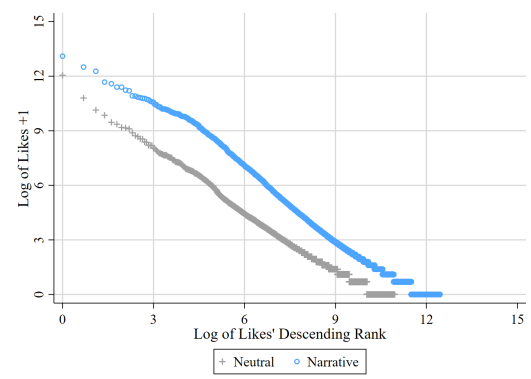
We begin by comparing tweets that feature a political narrative with those that do not in a descriptive way. Figure E.7a shows the Log-Log Rank Distribution of likes, separating tweets that include at least one character-role from those that only present characters in a neutral form. The x-axis reports 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. We proceed with reproducing the main results of the paper in Figure E.7b. The results 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.

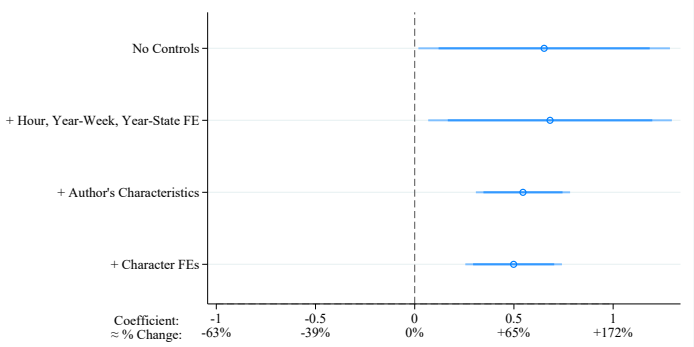
In sum, popularity exhibits the same narrative premium as virality, with effects that are stable across specifications. Given this close correspondence, we treat likes as a robustness check rather than a separate result.

Figure E.7: ***Impact of Political Narratives on Popularity of Tweets***

(a) Popularity of Political Narratives in Relevant Tweets (United States, 2010-2021)



(b) Regression Results-Impact of Political Narratives on Popularity (Likes)



Notes: The figure on the left 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 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. The figure on the right 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 (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, year × calendar-week, 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.

E.8 Impact of Political Narratives on Conversation

In this section of Appendix E.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 tweet receives.

We begin by exploring the distribution of comments in tweets featuring a political narrative and in tweets featuring characters only in neutral framing, in Figure E.8a. The figure 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.

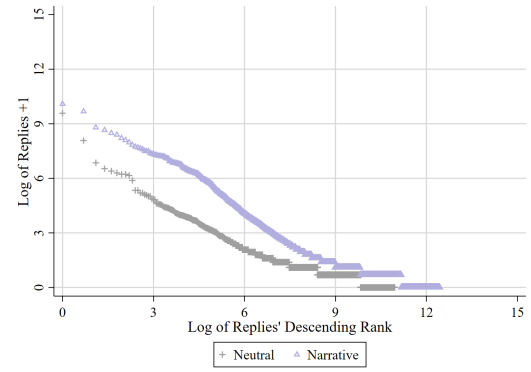
The distribution reveals several interesting insights. First, the distribution of comments appears to be highly skewed, consistent with what we observe for retweets. The linear shape of the log-log rank distribution suggests that the data generation process follows an exponential structure, with most 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.

We move from descriptive to regression analysis by reproducing the results of Figure 7 for conversation. Figure E.8b 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: $\approx e^{\beta} - 1$.

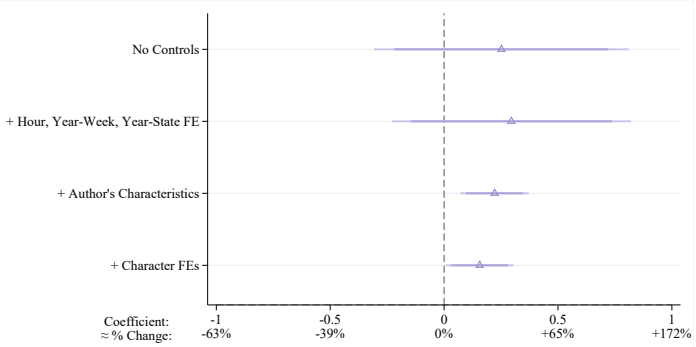
The results are mostly consistent with those found for virality in the paper, and popularity in the appendix. Although the first two models have large confidence intervals – suggesting noisier measures compared to retweets and likes – the overall effect indicates that narratives spark more conversation 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 E.8: ***Impact of Political Narratives on Conversation about Tweets***

(a) Conversation on Political Narratives in Relevant Tweets (United States, 2010-2021)



(b) Regression Results-Impact of Political Narratives on Conversation



Notes: The figure shows the log–log rank distribution of replies for relevant tweets, distinguishing between those with at least one character-role (purple) 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. 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, year × calendar-week, 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.

E.9 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, $\text{asinh}(Y)$ — and compute the corresponding Altonji-Elder-Taber (AET) ratios (Altonji, Elder, and Taber 2005).[25]

The AET ratio captures how much more strongly unobserved confounders would need to be correlated with both the treatment and the outcome, relative to the observed controls, in order to fully explain away the estimated effect. A ratio greater than one indicates that selection on un-

[25] We compute AET ratios for OLS-based specifications only. The AET framework relies on comparing how coefficients change as controls are added, a decomposition that depends on the linear structure of the estimator. In PPML, the conditional mean is modeled as $\exp(X\beta)$, and no straightforward analogue of the ratio exists under this non-linear specification. We choose the two transformations precisely to allow application of the AET framework while remaining close in spirit to our preferred PPML specification.

observables would need to exceed selection on observables; ratios above two or three are generally considered implausibly large. We choose these transformations to align with Subsection E.6, ensuring that the results are not driven by the particular functional form of the outcome and providing 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 close agreement between the two transformations further suggests that this conclusion is not an artifact of functional form. Taken together, these results provide strong evidence that the positive association between political narratives and tweet virality is not driven by omitted-variable bias: only implausibly strong and systematic selection on unobservables could eliminate the observed effect.

Table E.1: ***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	0.175	0.164	0.172
Observations	310,229	310,229	310,229	310,229
AET Ratio		4.40		4.20

Notes: The table reports OLS regression estimates of the effect of character-roles on tweet engagement. The dependent variables in Columns (1) and (2) are log-transformed retweet counts; Columns (3) and (4) use the inverse hyperbolic sine transformation instead, which accommodates zero-valued counts. The SHORT models include fixed effects for hour, year × calendar-week, and year × state, alongside controls for author characteristics and character fixed effects. The FULL models additionally control for language metrics, valence, and emotions. The Altonji-Elder-Taber (AET) ratio (Altonji, Elder, and Taber 2005) is a test for selection on unobservables: values above 1 are generally interpreted as evidence against omitted variable bias driving the results; our estimated AET ratio of 4 suggests that unobservable selection would need to be four times as large as observable selection to nullify the estimated effects. Standard errors are clustered at the author level.

E.10 Author Fixed Effects

This section addresses a central limitation of our observational analysis. The results in the paper are based on repeated cross-sections of tweets rather than a panel that follows the same users over time. For this reason, our estimates should be interpreted descriptively. In particular, we cannot exclude that some of the patterns we document reflect time trends, shifts in the platform environment, or other unobserved factors that move jointly with narrative usage and virality outcomes.

A natural way to strengthen the interpretation would be to exploit within-user variation using a panel of users and include user fixed effects. Doing so would clear out all time-invariant differences across authors, such as baseline popularity, writing style, or stable political orientation, and would focus the analysis on changes in narrative content within the same user. We do not observe a full user panel in our data. However, we do observe that some users appear more than once, because they authored multiple tweets that enter our dataset. We use this subset of repeat authors to conduct a within-sample check: do the estimated narrative premium and its magnitude change meaningfully when we absorb author-level heterogeneity with user fixed effects?

Table E.2 reports some results tackling this issue. Column (1) estimates a baseline relationship between narratives and virality for the subsample of users who authored more than one tweet in our dataset. Column (2) repeats the analysis while controlling for the same set of author characteristics used in the main specifications. Column (3) replaces these author controls with user fixed effects, thus absorbing all time-invariant user characteristics and relying only on within-user variation across tweets.

The baseline estimate in column (1) is not statistically significant in this restricted sample. We do not view this as evidence against the main results. The subsample is selective, and our dataset does not contain all tweets written by these users, but only those captured by our collection strategy. As a consequence, statistical power is lower and the remaining variation may be less representative of users' typical posting behavior. The more informative comparison is between columns (2) and (3). Focusing on the point estimates, the specification with observed author controls yields results that are qualitatively similar to the specification with user fixed effects. This pattern is reassuring: it suggests that the author controls used throughout the paper capture a substantial share of the across-user heterogeneity that user fixed effects would otherwise absorb.

Table E.2: ***Robustness Check – Author Fixed Effects***

Dependent Variable	**Retweets Count**		
	No FE	**+ Ctrls**	**FE**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Narratives	0.470	0.268	0.271
	(0.318)	(0.268)	(0.240)
	[0.140]	[0.318]	[0.260]
Sample (authors)	>1 tweets	>1 tweets	>1 tweets
Author FE			✓
Author controls		✓	
Mean Outcome	7.20	7.20	7.20
Pseudo R^2	0.00	0.59	0.85
Observations	84820	84820	84820

Notes: The table reports the results of our analysis of the virality of Political Narratives, accounting for author characteristics and author fixed effects. The models include only authors from whom we have at least 2 tweets in our sample. Each column corresponds to a different specification. Column (1), *No FE*, includes only the indicator for Political Narratives and does not control for any additional covariates. Column (2), *+ Ctrls*, adds author-level characteristics. Column (3), *FE*, replaces these controls with author fixed effects, hence exploiting within-author variation. Author characteristics include verified status, number of followers and followings, total number of tweets created, party affiliation, religiosity, higher education, and parenthood. The sample is restricted to tweets posted by authors with more than one tweet in the dataset who exhibit within-author variation, that is, authors who post both neutral and political narrative tweets.tweets.

This exercise is not a definitive solution to the identification limits of repeated cross-sectional data. User fixed effects address time-invariant author heterogeneity, but they do not eliminate time-varying confounders or broader platform-level trends that may affect both narrative content and engagement. We therefore treat this analysis as an internal robustness check rather than a causal design. At the same time, it helps motivate our first experiment. In a controlled survey setting, we test whether narratives generate a premium in engagement even when exposure is randomized. The experimental evidence supports the main conclusion: narratives carry a virality premium even under experimental control.

F Political Narratives Correlation with Survey Data: Cooperative Election Study

This section examines whether the prevalence of political narratives in public discourse correlates with public policy preferences in nationally representative survey data. Using the Cooperative Election Study (CES) for the period 2014–2021, we relate the state-year share of Green Tech and Regulations narratives cast as hero or villain to respondents' policy preferences in the same state

and year. We interpret these correlations descriptively: the relationship could reflect narratives shaping preferences, preferences shaping the narratives users choose to post, or both.

Table F.3 reports results for renewable energy preferences. The dependent variable is an indicator for supporting a minimum renewable fuel requirement for states, even at the cost of slightly higher electricity prices. The key independent variables are the state-year shares of Green Tech cast as hero or villain. We find that the hero share is positively but not significantly associated with pro-renewable preferences across specifications. In contrast, the villain share is strongly and consistently negatively correlated with support for renewable energy.

Table F.4 presents analogous results for Regulations narratives and public support for the Environmental Protection Agency (EPA). The dependent variable is an index averaging support for granting the EPA authority to regulate carbon dioxide emissions and for strengthening enforcement of the Clean Air Act and the Clean Water Act. Again, hero framing of Regulations is not significantly associated with EPA support. However, villain framing shows a negative and statistically significant association in the most restrictive specification with state fixed effects, with a coefficient of -0.244.

Taken together, the results across both policy domains point to an asymmetric pattern: villain framing is more strongly and consistently associated with policy attitudes than hero framing, a finding that mirrors the broader result in the paper that villain narratives tend to be particularly salient in climate policy discourse.

Table F.3: ***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	0.019	0.019	0.018			
Tech-Hero	(0.018)	(0.018)	(0.018)			
	[0.283]	[0.283]	[0.322]			
Share of Green				-0.489	-0.489	-0.489
Tech-Villain				(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 F.4: ***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	-0.066	-0.066	-0.064			
Regulations-Hero	(0.056)	(0.056)	(0.055)			
	[0.243]	[0.243]	[0.242]			
Share of				-0.088	-0.088	-0.244
Regulations-Villain				(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.

G Political Narratives in Other Media

The main analysis investigates the virality of narratives on Twitter/X. This appendix examines whether similar patterns in narrative use extend beyond social media, by applying our pipeline to two additional media sources: newspapers and television. Although we cannot reproduce the virality results in these settings, finding similar narrative distributions across media provides external validation of the framework and suggests that the patterns we document reflect broader tendencies in climate policy discourse rather than features specific to Twitter/X.

G.1 Newspapers

In preparing and classifying newspaper articles, we follow the same procedure used for tweets as closely as possible. We source a random and representative sample from the three most widely circulated newspapers in the US: The New York Times, The Wall Street Journal, and USA Today.

Articles are downloaded from Factiva in four-year intervals covering 2010–2021. For each newspaper, we download the 333 or 334 most popular articles from each interval (2010–2013, 2014–2017, 2018–2021), yielding 3,000 articles per outlet. We apply the same keyword filter used for tweets (Oehl, Schaffer, and Bernauer 2017) (see Section A.1) and minimally adapt the classification prompts, replacing references to tweets with references to newspaper articles.

The distribution of character-roles in newspaper articles closely mirrors the patterns observed on Twitter. Among human characters, the three most frequent character-roles are US Democrats-Hero (7.52%), Corporations-Villain (6.33%), and US Republicans-Villain (6.14%). Among instrument characters, Fossil Industry-Villain is the most prevalent (14.98%), followed by Green Tech-Hero (9.92%) and Regulations-Hero (roughly 6%). These rankings are broadly consistent with those observed on Twitter, with villain framing of Fossil Industry and hero framing of Green Tech dominating in both settings.

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

(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

(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.

G.2 Television

In preparing and classifying the television transcripts, we follow as closely as possible the procedure used for tweets and newspaper articles. We construct a dataset of television news content by collecting transcripts from two major U.S. cable news networks: Fox News and MSNBC. These outlets were selected because they represent two of the most widely viewed and politically influential television news channels during the period of study. We retrieve transcripts using the GDELT Television API, focusing on episodes aired between 2010 and 2020 that mention the keywords "climate change" or "global warming". Because full television transcripts are typically long and exceed the token limits of language models, we divide each transcript into smaller, consecutive text windows suitable for classification.

Table G.2 reports the distribution of character-roles in relevant television transcripts. Similar to the patterns observed on Twitter and in newspapers, we find that political narratives are prevalent in televised news coverage. In particular, Fossil Fuels frequently appear in Villain roles, while Green Technologies and Regulations are often framed as Heroes. Human actors such as US Democrats, US Republicans, and Corporations also appear in clearly defined narrative roles.

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

(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

(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.

H Additional Information on the Survey Experiments

This appendix provides additional information on the experimental designs adopted in our online surveys. We begin with the *virality experiment*, which consists of a single survey wave, before turning to the *persuasion experiments* – a set of three separate experiments with connected follow-up surveys. The final part of the appendix presents balance checks for the samples obtained through randomization in each experiment.

H.1 Virality Experiment

The *virality experiment* is a single-wave survey experiment designed to validate the main observational finding of the paper: that narratives carry a virality premium relative to neutral versions of the same factual content. We implement the experiment via Prolific and Qualtrics.

We recruit 600 participants, sampled to be representative of the US population along key demographics and political affiliation. Participants are told that the survey studies the spread of content in social interactions. They read six short text extracts and, after each extract, make a single binary decision: whether to share the text with another real Prolific user. Participants are informed that the pairing is real, that their identity remains anonymous, and that the recipient is drawn from a representative US sample. To ensure data quality, participants completed two comprehension checks before proceeding; those who failed both were asked to leave the survey. The final sample of 600 participants thus consists entirely of respondents who passed at least one comprehension check.

Each extract is drawn from one of six paired texts: a neutral version and a narrative version. Within each pair, the two versions contain the same factual content and mention the same characters, differing only in role assignment: the narrative version casts characters as hero, villain, or victim, whereas the neutral version mentions them without assigning any role. Randomization occurs along two dimensions. First, for each pair, participants are randomly assigned to view either the narrative or the neutral version. Second, the order of the six pairs is randomized.

Table H.1 and Table H.2 report the full set of six text pairs. For each text we list the pair number, whether the text is neutral or narrative, and — for the narrative version — the character-role assignments; in the neutral version, characters appear in neutral form only. The pair number serves purely as a label, given that the order is randomized in the experiment.

This design yields six sharing decisions per participant, increasing statistical power and allowing us to estimate the effect of narrative framing by comparing sharing rates under randomized exposure, holding constant underlying content, and controlling for participant fixed effects — hence exploiting within-participant variation. After completing the six decisions, participants answer a short set of background questions, including self-reported political leaning and measures of social media use.

We provide access to the Qualtrics survey and the pre-registration at the links below:

- Qualtrics survey

- Pre-Reg

Table H.1: ***Texts Used in the Virality Experiment (i.)***

Pair	Type	Text	Character-Roles
1	Neutral	Organ donation and transplantation involve donors, donor families, transplant surgeons, and multidisciplinary medical teams. According to the Health Resources and Services Administration, there were 46,632 organ transplants performed in the United States in 2023. Donated organs are recovered according to medical and legal criteria, then matched to recipients through established allocation systems. Transplant surgeons perform the operations, while anaesthetists, nurses, and other specialists manage perioperative and post operative care. Outcomes depend on organ quality, timing, patient health, and long term clinical follow up. Policy and clinical discussions focus on allocation rules, waiting lists, and how to increase the number of usable organs.	People (Neutral); Medical Staff (Neutral)
1	Narrative	Organ donors and transplant surgeons quietly open doors that would otherwise stay shut. In 2023, there were 46,632 organ transplants in the United States, according to the Health Resources and Services Administration. Each one began with a donor or a family, often in the middle of grief, choosing to let an organ go to someone they would never meet. Surgeons and their teams then turned that choice into a working heart, kidney, or liver for a patient who had been waiting, sometimes for years. For many of those patients, including children, it meant trading a calendar of hospital visits for the chance to think about school, work, or the next holiday instead. It is a rare kind of partnership where loss and skill add up to more ordinary days, and it is easy to forget how often that happens.	People (Hero); Medical Staff (Hero)
2	Neutral	The United States has 63 congressionally designated national parks, managed by the National Park Service, covering about 52 million acres in 30 states and two US territories and including landscapes such as mountains, canyons, forests and coastal areas. The parks are governed by federal laws and agency regulations. Park rangers and other staff monitor conditions, maintain trails and facilities, provide information and respond to incidents under established procedures. Funding comes from federal appropriations, user fees and other revenue sources and the system is subject to periodic planning and review.	National Parks (Neutral); Park Staff (Neutral)
2	Narrative	American national parks and the people who care for them quietly protect some of the most remarkable ground in the country. From high glaciers and red rock canyons to old growth forests and coastline, 63 national parks stretch across about 52 million acres in 30 states and two US territories and still feel wild when you step into them. Rangers and park staff walk the trails in every season, clear fallen trees, repair signs and watch over wildlife to ensure open views and clear paths instead of closures and warning tape. The parks offer something simple, a place where long weeks and crowded streets can be traded for fresh air, distance and the steady rhythm of walking, a rare thing in the world and one of the country's quiet achievements.	National Parks (Hero); Park Staff (Hero)
3	Neutral	Health insurance in the United States involves regular premium payments for coverage under private and public plans, and Census Bureau estimates suggest that in 2023 health plans covered about 305.2 million people at some point in the year. In employer sponsored and individual coverage insurers apply written benefit rules to incoming claims to determine which services are covered and how costs are split between the plan and the member through deductibles, co payments and other forms of cost sharing. We should think more about how to to calibrate benefit limits and administrative procedures to balance customer reliability and insurance requirements.	Insurance Companies (Neutral); Regulations (Neutral)
3	Narrative	In the United States the Census Bureau estimates suggest that in 2023 health plans covered about 305.2 million people. Families hand over part of every paycheck for coverage that is supposed to stand between them and huge medical bills. However, when a serious illness or injury hits, they discover that insurers still decide which parts of the hospital or specialist charges count as covered and which costs come back to them. Plan rules, networks and prior authorizations can leave them owing far more than they ever expected at the exact moment they feel least able to fight.	Insurance Companies (Villain); Regulations (Villain)
4	Neutral	Pharmaceutical companies developed and marketed opioid painkillers that have been widely used to treat severe and chronic pain, particularly in acute care, cancer treatment, and end of life settings. At the same time, expanded prescribing raised concerns about dependence and overdose risks. According to the CDC, there were an estimated 54,743 opioid involved overdose deaths in 2024. Policy discussions continue about how to balance the medical benefits of opioids with safeguards around prescribing, product design, and public health responses.	Big Pharma (Neutral); People (Neutral)
4	Narrative	Big pharma didn't just sell painkillers — it sold a story, pushed prescriptions, and kept the money flowing even as dependence spread through ordinary families. Years later, the damage is still showing up in real life: the CDC estimates 54,743 opioid-involved overdose deaths in 2024. The companies that created the worst drug related issue in history are still cashing in profits while Americans are the ones left living with the fallout in their homes and communities.	Big Pharma (Villain); People (Victim)

Notes: The table reports four of the six pairs of texts used in our Virality Experiment. Each pair is constituted by a neutral text and its narrative correspondent. The two texts share the same characters and hold the factual information provided constant. The remaining two pairs are provide below.

Table H.2: ***Texts Used in the Virality Experiment (ii.)***

Pair	Type	Text	Character-Roles
5	Neutral	Parking enforcement officers and tow truck drivers both work in systems that manage vehicle parking and movement. One recent estimate suggests there are about 5,675 parking enforcement officers employed in the United States. Enforcement officers issue citations or warnings according to local ordinances and posted regulations. Tow truck drivers respond to calls from law enforcement, property owners, or roadside assistance services to move or recover vehicles. Their work is guided by legal rules, contracts, and dispatch priorities.	Officers (Neutral); Tow Truck Drivers (Neutral)
5	Narrative	Parking enforcement officers and tow truck drivers rarely meet people on good days. One recent estimate suggests there are about 5,675 parking enforcement officers employed in the United States, so each officer handles many tense encounters in a typical week. Most interactions start with someone spotting a ticket or realizing their car has been moved, and the frustration lands on whoever is standing there in a uniform or reflective vest. Both jobs run on routes, rules, and dispatch instructions they did not write, with little room to change outcomes on the spot. They step into arguments all day while mostly just trying to finish their rounds and clear the list in front of them.	Officers (Victim); Tow Truck Drivers (Victim)
6	Neutral	Large technology platforms host billions of user accounts and process high volumes of content and data. A recent global report estimates there were about 5.04 billion active social media user identities worldwide at the start of 2024, equal to roughly 62 percent of the world's population. Companies employ internal teams to develop policies on safety, privacy, and acceptable content, and trust and safety staff apply these policies to specific cases using established guidelines and tools. When policies change, users, advertisers, and external observers react in different ways. Public discussions focus on transparency, consistency, and how platforms respond to competing expectations.	Big Tech Workers (Neutral); Users (Neutral) –
6	Narrative	People talk about Big Tech as if it were a single mind, but inside each company are teams that spend every day being blamed from all sides. A single change in a feed or policy can trigger millions of complaints, press stories, and angry posts, even if the change was meant to fix another problem. A recent global report estimates about 5.04 billion active social media user identities worldwide at the start of 2024, so even small tweaks to how platforms work can affect huge numbers of people at once. Trust and safety staff spend long days reviewing bad content and impossible edge cases, knowing that any decision can be criticized as too weak or too strict. Both the platforms and the people inside them live with constant pressure to get it perfect in an environment where everyone only sees what went wrong.	Big Tech Workers (Victim); Users (Victim)

Notes: The table reports two of the six pairs of texts used in our Virality Experiment. Each pair is constituted by a neutral text and its narrative correspondent. The two texts share the same characters and hold the factual information provided constant. The remaining four pairs are provide above.

H.2 Persuasion Experiments

The second set of experimental evidence concerns the persuasive power of viral and frequently used narratives. We tackle the following question: does the virality of narratives translate into greater persuasiveness? The question is important because, although we provide evidence that narratives carry a virality premium, it remains unclear whether they are also more effective in changing people's beliefs, preferences, and behavior. We address this through three separate experiments conducted via Prolific and Qualtrics.

We run each experiment with approximately 1,000 participants, drawn from a representative sample of the US population based on age, ethnicity, gender, and political affiliation. Unlike the virality experiment, we do not automatically screen out participants who fail comprehension checks. However, we pre-registered the requirement that each experiment yield at least 800 usable observations, and additional participants were recruited as needed to meet this threshold. Within each experiment, participants are randomly assigned to either a control or a treatment group. Both

groups are told that the survey studies the impact of exposure to social media content, and are shown a fictitious feed of three posts resembling Twitter or Facebook. Two posts are identical across groups and serve as obfuscation. The third post differs: in the control group it conveys factual information and depicts characters in a neutral role, while in the treatment group it presents the same information but casts characters in drama triangle roles.

In each experiment, the treatment post features different character-roles, with the exception of Green Tech-Hero, which appears in all three designs to ensure comparability across experiments. We label each experiment according to the character-roles featured in the treatment post. 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 all three experiments, assessing whether they recalled the factual information they had read and whether they remembered the content of the post.

The character-role combinations featured in the treatment posts were informed by the results of the observational analysis. Specifically, we use narrative structures — such as Green Tech-Hero and US People-Hero — that emerged as particularly frequent or viral in our Twitter data. Table H.3 reports the three neutral-narrative text pairs. The first pair was used in the Hero-Hero experiment, the second in the Hero-Hero-Villain experiment, and the third in the Villain-Villain-Hero experiment.

In the main sections of the paper, we provide further details on design choices. We provide access to the experiments through the links below:

- Main Experiment-Hero-hero: Qualtrics survey
- Main Experiment-Hero-Hero-Villain: Qualtrics survey
- Main Experiment-Villain-Villain-Hero: Qualtrics survey
- Follow-up Experiment (all conditions): Qualtrics survey

For transparency, we also provide the link to the pre-registrations of each experiment at the links below:

- Main Experiment-Hero-Hero: Pre-Reg
- Follow-Up-Hero-Hero: Pre-Reg
- Main Experiment-Hero-Hero-Villain: Pre-Reg
- Follow-Up-Hero-Hero-Villain: Pre-Reg
- Main Experiment-Villain-Villain-Hero: Pre-Reg

- Follow-Up-Villain-Villain-Hero: Pre-Reg

Table H.3: ***Texts Used in the Persuasion Experiments***

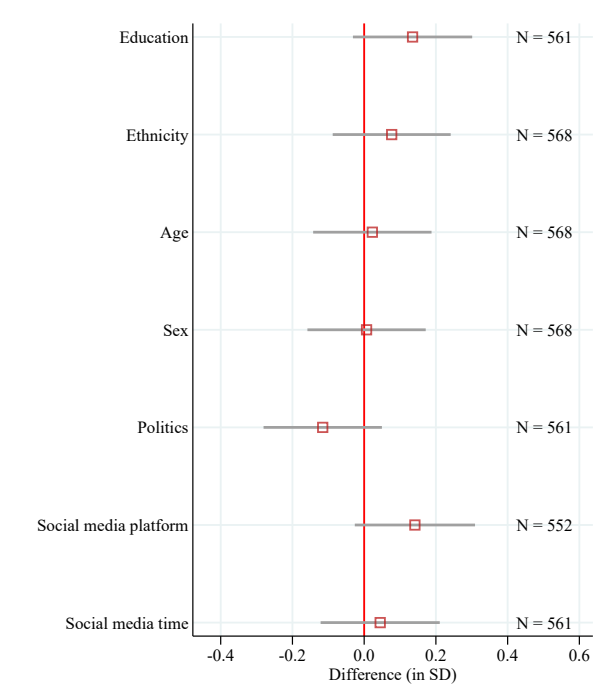
Experiment	Type	Text	Character-Roles
Hero-Hero	Neutral	Climate change policies remain a hotly debated political issue in the U.S. As of 2023, solar energy has provided 163 billion kWh of electricity. Solar provides clean energy, but concerns about overall reliability are key aspects that might prevent more regular Americans from installing solar panels on their homes. Besides solar, there are also other green technologies like wind or geothermal. The future will show if these new energy sources can become a key part of US energy production. #GreenTech #Energyproduction	Green Tech (Neutral); The People (Neutral)
Hero-Hero	Narrative (Hero-Hero)	Across the U.S., families are helping secure energy independence by adopting solar energy, already generating an impressive 163 billion kWh of electricity in 2023. Solar energy is decentralizing power, bringing reliable, clean electricity directly to homes and reducing emissions. With constant innovation improving panels' efficiency, green tech is empowering everyday Americans to build a more independent and sustainable future. This shows that change is not only possible, it's happening. #CleanEnergy #SolarFuture	Green Tech (Hero); The People (Hero)
Hero-Hero-Villain	Neutral	Climate change policies remain a hotly debated political issue in the U.S. Some people favor green technology like solar, wind or geothermal. As of 2023, solar energy has provided 163 billion kWh of electricity. However, others prefer to continue relying on fossil resources like oil, gas and coal. The pros and cons of fossil energy sources versus renewables remain key points of debate. Some argue that policies and more regulation are necessary to help renewables gain a higher market share, while others emphasize challenges of regulating too much. #ClimateDebate #Energyproduction	Green Tech (Neutral); Regulations (Neutral); Fossil Industry (Neutral)
Hero-Hero-Villain	Narrative (Hero-Hero-Villain)	Fossil fuel companies still try to lobby behind closed doors in desperate attempts to block innovation and hide their environmental and health toll. But the tide is turning: solar power alone already produced an impressive 163 billion kWh of clean electricity in 2023, contributing to cleaner air and energy independence. Strong regulations requiring transparency on environmental risks and fast-tracking renewables empower green tech to lead us to a healthier, more sustainable future. They help us to choose progress over pollution. #ClimateDebate #Energyproduction	Green Tech (Hero); Regulations (Hero); Fossil Industry (Villain)
Villain-Villain-Hero	Neutral	Climate change policies remain a significant and debated issue in U.S. politics. Currently, one source of U.S. energy production is still fossil fuels, with the public discussing their future role and impact. At the same time, alternative energy sources such as solar have become part of the energy landscape. Solar contributed 163 billion kWh to electricity generation in 2023. Against this background, it seems important to carefully evaluate the efficiency, costs, and implications of all these energy options.	Fossil Industry (Neutral); Corporations (Neutral); Green Tech (Neutral)
Villain-Villain-Hero	Narrative (Villain-Villain-Hero)	The energy system isn't broken, it's rigged. Fossil fuel corporations, driven by profit at any cost, control markets and bury competition before it can grow. Meanwhile, fossil energy endangers our health with toxic air pollution leading to rising rates of respiratory illness. Solar energy is clean and produced 163 billion kWh in 2023 - thanks to innovation - but it's a fraction of what it could be if markets weren't designed to favor polluters. Fossil fuels don't just dominate, they suffocate progress while endangering our planet and our lives. #ClimateDebate #Energyproduction	Fossil Industry (Villain); Corporations (Villain); Green Tech (Hero)

Notes: The table reports the three pairs of texts used in our *persuasion experiments*. Each pair is constituted by a neutral text and its narrative correspondent. The two texts share the same characters and hold the factual information provided constant. The remaining four pairs are provide above.

H.3 Balance Tests

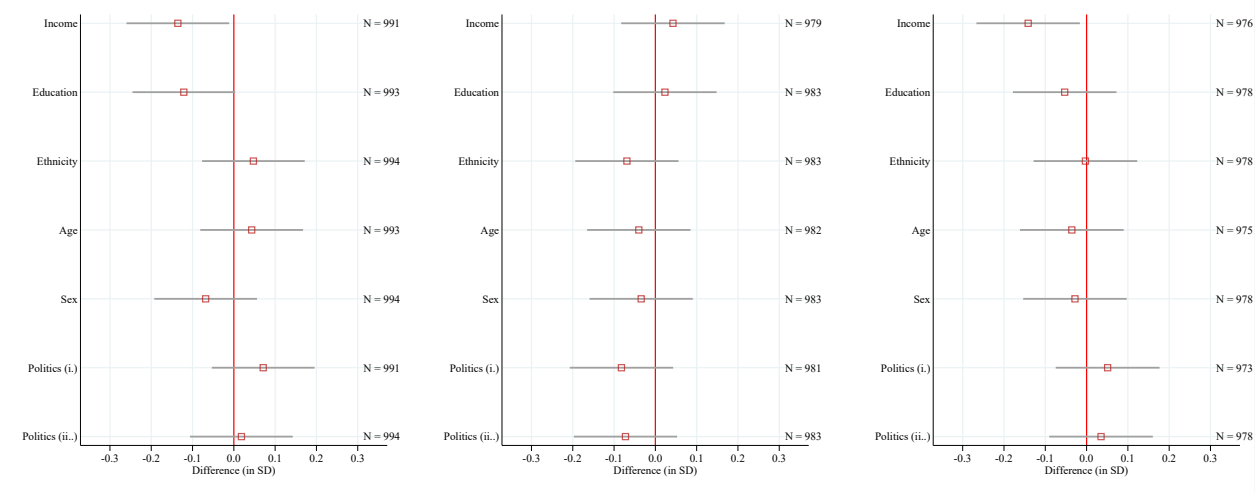
In this section we provide evidence of the balance of samples obtained through the randomization process applied for each experimental survey we conducted. In particular, Figure H.1 provides balance tests for several individual characteristics of participants in the treatment and control group from our *virality experiment*. Figure H.2 shows the same for the three surveys part of our *persuasion experiments*. For all figures, we obtain the coefficients by regressing each individual feature on an indicator of assignment to the treatment group.

Figure H.1: ***Balance of Individual Characteristics of Participants to the Virality Experiment***



Notes: The figure shows control-treatment balance tests for several individual characteristics among the participants of the first experiment, testing the virality of narratives. The 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.

Figure H.2: ***Balance of Individual Characteristics of Participants to the Persuasion Experiments***



(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.

H.4 Attrition of Follow-Ups

All three surveys in our *persuasion experiments* are organized to be conducted in one main experimental day and to have then a follow up survey, the day after. In Table H.4 we show the attrition rate for each experimental wave: HH, HHV, and VVH.

Table H.4: ***Attrition Rates by Experiment Wave***

Wave	Experiment Day	Follow Up	Attrition Rate
HH	994	792	20.3%
HHV	983	804	18.2%
VVH	978	701	28.3%

Notes: The table reports the number of participants on the day of the experiment, the number who returned for the follow-up survey one day later, and the implied attrition rate for each of the three persuasion experiment waves (Hero-Hero, Hero-Hero-Villain, Villain-Villain-Hero).

I Regression Tables Underlying the Main Figures: Experimental Data

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 7. Table I.1 provides the models for the Paper Figure 12a, Table I.2 for the Paper Figure 12b, Table I.3 about

the donation results in Figure 13. Table I.4 and Table I.5 provide additional details on our memory results, from the Paper Figure 14a and Figure 14b.

Table I.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 12a in the paper is the reference plot.

Table I.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 12b in the paper is the reference plot.

Table I.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 13 in the paper is the reference plot.

Table I.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 14a are the reference plots.

Table I.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 14b in the paper is the reference plot.

J Heterogeneity of Results from the Virality Experiment

In the main paper we provide evidence of the effect of narratives on sharing decisions taken by the participants of our *virality experiment*. Both villain and hero narratives are shared more than neutral content. The two roles drive the overall effect, while victim narratives show a null effect, leading to a general positive impact of narratives on sharing decisions. In this appendix, we explore heterogeneous effects of this result by political leaning, social media platform used, and time spent on social media daily. Table J.1, Table J.2, and Table J.3 reproduce the baseline specification from Table 3, showing the impact of narratives on sharing decisions separately for each of these dimensions, respectively.

The results in Table J.1 provide a clear picture. When exploring the impact of narratives across participants self-identifying as Democrats — Column (1) — Republicans — Column (2) — and Other/Independent — Column (3) — a consistent pattern emerges: the narrative premium on sharing is virtually the same across the political spectrum. This addresses a concern raised by the observational data analysis, namely that Twitter users may constitute a self-selected and politically unrepresentative sample of the population, skewed towards Democratic views during the years of our analysis. These results seem to reassure that the results are consistent and independent of political views.

Table J.2 addresses a related concern: Twitter users may also be a population particularly prone

to using and engaging with political narratives. We therefore reproduce the main result separately by participants' self-reported social media platform use. Column (1) includes participants using Twitter, Facebook, and Instagram; Column (2) Twitter only; Column (3) Facebook only; and Column (4) Instagram only. The results are fairly consistent – virtually identical for Twitter and Facebook, and only slightly weaker for Instagram users. This suggests that the narrative premium on sharing is not specific to Twitter.

Finally, in Table J.3 we also show that results are consistent across self-reported time spent on social media. Independently to the time spent on social media, narratives seem to have a positive effect on the decision of sharing the text or not sharing the text. Noticeably, the strongest effect appears to be the one for the sub-set of population that spends three or more hours on social media platforms.

Table J.1: ***Experimental Evidence on the Virality of Political Narratives: Heterogeneity by Political Leaning***

Dependent Variable	**Decision to Share Content**		
	Participants by Politics:		
	Democratic	**Republican**	**Other**
	Coeff./SE/p-value		
	(1)	(2)	(3)
Treatment	0.169	0.128	0.168
	(0.030)	(0.034)	(0.028)
	[0.000]	[0.000]	[0.000]
Participant FE	✓	✓	✓
Sample: All Narratives	✓	✓	✓
Mean Outcome Control Group	0.39	0.44	0.35
Observations	1158	930	1380

Notes: The table reports OLS estimates from models estimating the impact of narratives on virality. The data is collected through our first experiment, where 600 participants were asked to read six short text extracts and choose whether they would share the text with another Prolific participant. The pairing with the receiving participant is real and the choice is a real stake outcome. For each choice participants were randomly assigned to see either the neutral or the narrative version of the same text. Neutral and narrative contain the same factual information and differ only in the presence of narrative roles. The data is organized in a panel where there are six entries per participant; the choices were in random order and we use participant FE to exploit within participant variation. The six pairs were divided in: two pairs focused on hero narratives, two on villains and two on victims. The different models show the effects for sub-samples of participants grouped by their self-reported political leaning.

Table J.2: ***Experimental Evidence on the Virality of Political Narratives: Heterogeneity by Social Media Platform Used***

Dependent Variable	**Decision to Share Content**			
	Subsample by Social Media Used:			
	Twitter,Facebook,Instagram	**Twitter**	**Facebook**	**Instagram**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Treatment	0.188	0.190	0.201	0.099
	(0.044)	(0.072)	(0.039)	(0.053)
	[0.000]	[0.012]	[0.000]	[0.068]
Participant FE	✓	✓	✓	✓
Sample: All Narratives	✓	✓	✓	✓
Mean Outcome Control Group	0.45	0.27	0.39	0.39
Observations	450	192	606	456

Notes: The table reports OLS estimates from models estimating the impact of narratives on virality. The data is collected through our first experiment, where 600 participants were asked to read six short text extracts and choose whether they would share the text with another Prolific participant. The pairing with the receiving participant is real and the choice is a real stake outcome. For each choice participants were randomly assigned to see either the neutral or the narrative version of the same text. Neutral and narrative contain the same factual information and differ only in the presence of narrative roles. The data is organized in a panel where there are six entries per participant; the choices were in random order and we use participant FE to exploit within participant variation. The six pairs were divided in: two pairs focused on hero narratives, two on villains and two on victims. The different models show the effects for sub-samples of participants grouped by what they self-report being the social media platform they use daily.

Table J.3: ***Experimental Evidence on the Virality of Political Narratives: Heterogeneity by Time Spent on Social Media Platforms***

Dependent Variable	**Decision to Share Content**			
	Subsample by Social Media Used:			
	< 1 Hour	**1-2 Hours**	**2-3 Hours**	**3+ Hours**
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Treatment	0.156	0.146	0.132	0.193
	(0.032)	(0.029)	(0.043)	(0.042)
	[0.000]	[0.000]	[0.003]	[0.000]
Participant FE	✓	✓	✓	✓
Sample: All Narratives	✓	✓	✓	✓
Mean Outcome Control Group	0.34	0.40	0.45	0.39
Observations	1212	1032	552	570

Notes: The table reports OLS estimates from models estimating the impact of narratives on virality. The data is collected through our first experiment, where 600 participants were asked to read six short text extracts and choose whether they would share the text with another Prolific participant. The pairing with the receiving participant is real and the choice is a real stake outcome. For each choice participants were randomly assigned to see either the neutral or the narrative version of the same text. Neutral and narrative contain the same factual information and differ only in the presence of narrative roles. The data is organized in a panel where there are six entries per participant; the choices were in random order and we use participant FE to exploit within participant variation. The six pairs were divided in: two pairs focused on hero narratives, two on villains and two on victims. The different models show the effects for sub-samples of participants grouped by their self-reported time spent on social media.

K Robustness Checks on the Virality Experiment

K.1 Randomization Inference: Virality Experiment

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 K.1: ***Experimental Evidence on the Virality of Political Narratives: Panel Dataset at Prolific Participant Level - Randomized Inference***

Dependent Variable	**Decision to Share Content**				
	All Types of Narratives		**Hero Narratives**	**Villain Narratives**	**Victim Narratives**
	Coeff./SE/p-value				
	(1)	(2)	(3)	(4)	(5)
Treatment	0.149	0.157	0.218	0.222	0.046
	(0.018)	(0.018)	(0.035)	(0.036)	(0.033)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.201]
Participant FE		✓	✓	✓	✓
Mean Outcome Control Group	0.39	0.39	0.45	0.45	0.27
Observations	3469	3468	1156	1156	1156

Notes: The table reports OLS estimates from models estimating the impact of narratives on virality. The data is collected through our first experiment, where 600 participants were asked to read six short text extracts and choose whether they would share the text with another Prolific participant. The pairing with the receiving participant is real and the choice is a real stake outcome. For each choice participants were randomly assigned to see either the neutral or the narrative version of the same text. Neutral and narrative contain the same factual information and differ only in the presence of narrative roles. The data is organized in a panel where there are six entries per participant; the choices were in random order and we use participant FE to exploit within participant variation. The six pairs were divided in: two pairs focused on hero narratives, two on villains and two on victims. Columns 1-4 report results for the full sample of participants while columns 5-7 subset participants by their self reported political stance. We compute the p-value using the STATA command Ritest for randomized inference, using 1000 repetitions, with strict and two-sided specification. Table 3 in the paper is the reference table.

L Robustness Checks on the Persuasion Experiments

L.1 Manipulation Check

In this section we provide evidence that the treatment manipulation in our *persuasion experiments* 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 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 L.1 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.

Table L.1: ***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	0.281	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 | 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.

L.2 Power and Minimum Detectable Effects

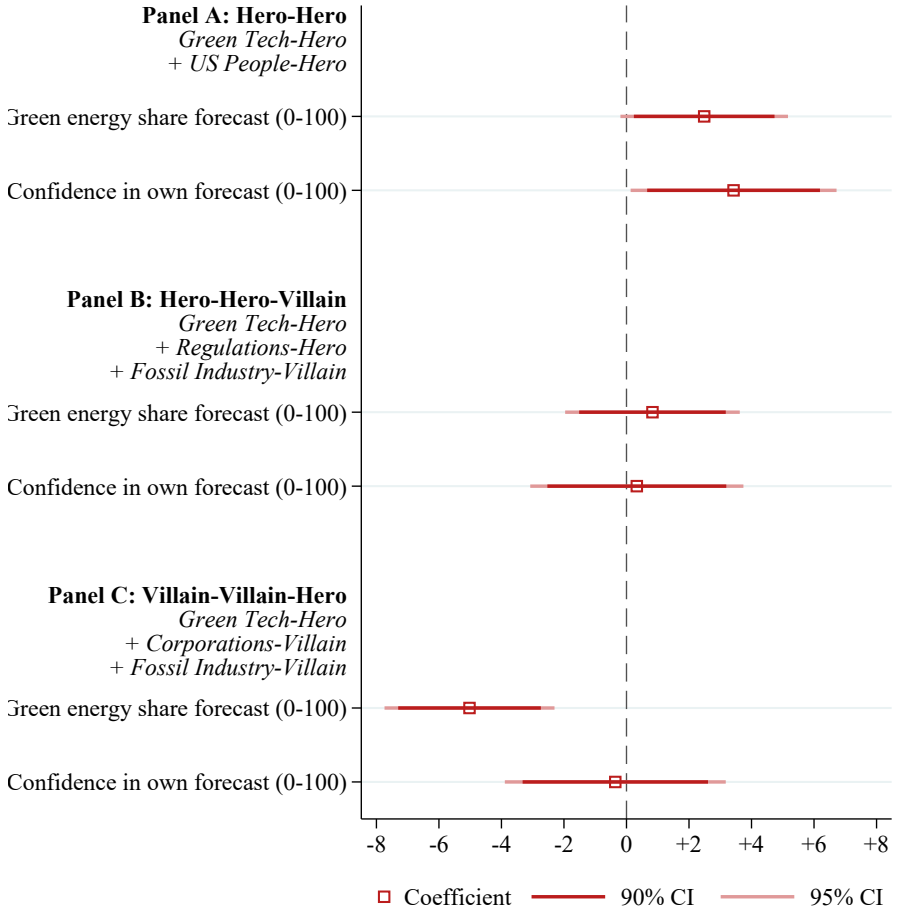
To assess whether our null result on policy preferences is informative or merely underpowered, we compute minimum detectable effect (MDE) sizes for each of the three persuasion experiments. For a two-sided test at $\alpha = 0.05$ and 80% power, the MDEs on the 7-point Likert policy-support scale are 0.267 Likert points in the Hero-Hero experiment, 0.255 in Hero-Hero-Villain, and 0.353 in Villain-Villain-Hero. In standardized units, these correspond to roughly 0.18 standard deviations in all three cases. The null on policy preferences should therefore be read as evidence against policy-support shifts of approximately 0.18 SD or larger; smaller shifts cannot be reliably distinguished from zero in our per-experiment samples.

L.3 Confidence in Beliefs

In this section we complement the results on belief change shown in Figure 12a. In particular, we show additional results on respondents' confidence in their own belief forecasts across all *persuasion experiments*. Figure L.1 reproduces the corresponding paper figure, adding a coefficient estimate for the effect of narratives on respondents' confidence in their own forecast. Confidence is measured on a scale from 0 to 100.

Narratives have virtually no effect on respondents' confidence in their own forecast. The sole exception is the Hero-Hero experiment, where the narrative has a weak but positive effect on confidence, increasing it by roughly 3 points. One interpretation is that the Hero-Hero narrative may have made respondents more optimistic about the future of green technologies, hence also more confident in their world view.

Figure L.1: ***Impact on Beliefs: Forecast for the Future***



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 12a, 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]%?'*.

L.4 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 L.2 and Table L.3 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 L.2: ***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 14a in the paper is the reference plot.

Table L.3: ***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

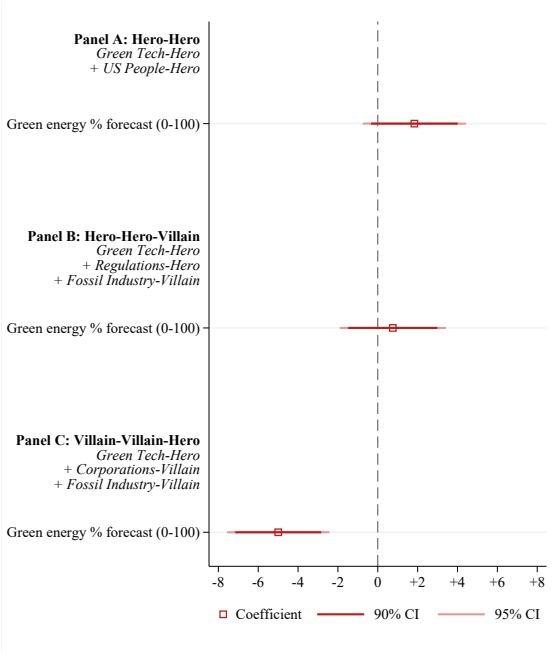
Notes: The table reports OLS regression results on the effect of narratives on participants' memory of the factual information embedded in both control and treatment tweets, measured on the day after of the experiment. The dependent variable comes from the question: "How many billion kWhs were generated 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 14a in the paper is the reference plot.

L.5 Output Excluding Controls

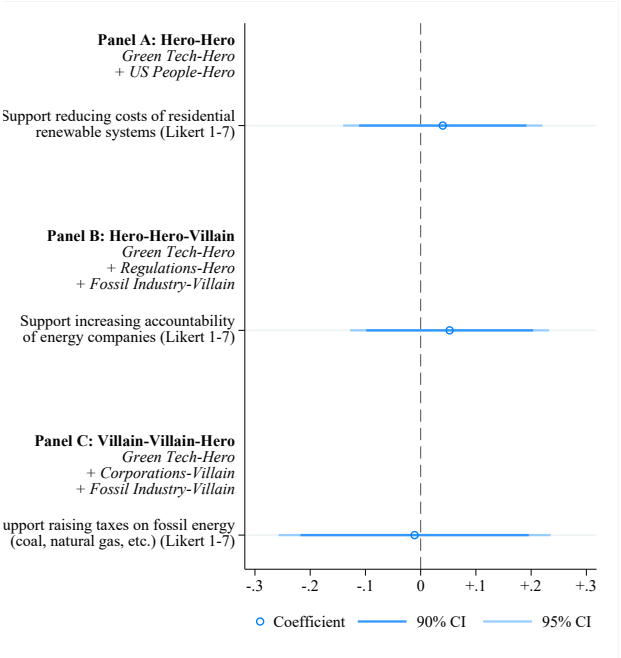
In this section we provide a robustness check on the model specifications used for our experimental results from the *persuasion experiments*. In particular, all the exhibits concerning our experimental results in the main paper include in the models' specification controls for the individual characteristics of the participants. In this section we report the main results excluding these controls.

Figure L.2: ***Experimental Results-Impact of Political Narratives on Beliefs and Policy Preferences (No Controls)***

(a) Impact on Beliefs: Forecast for the Future

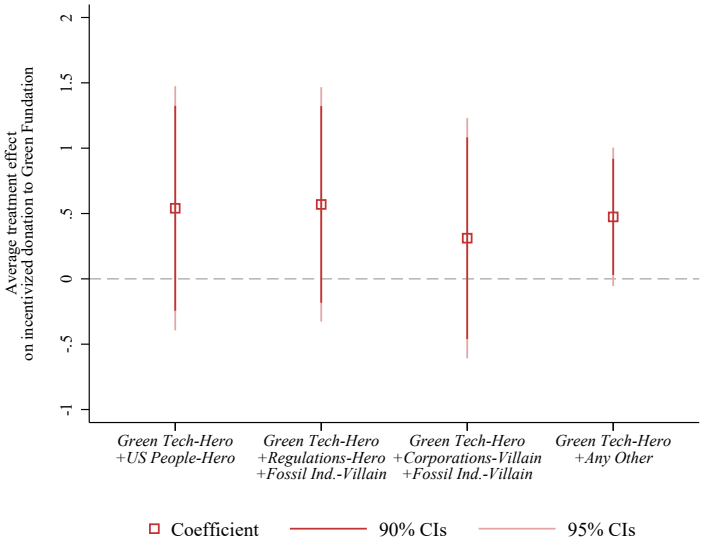


(b) Impact on Stated Policy Preferences: Policy Support



Notes: The figures reproduce results in the paper exhibit Figure 12 but excluding controls from all regression models.

Figure L.3: ***Experimental Results-Impact of Political Narratives on Revealed Preferences about Character GREEN TECH (No Controls)***

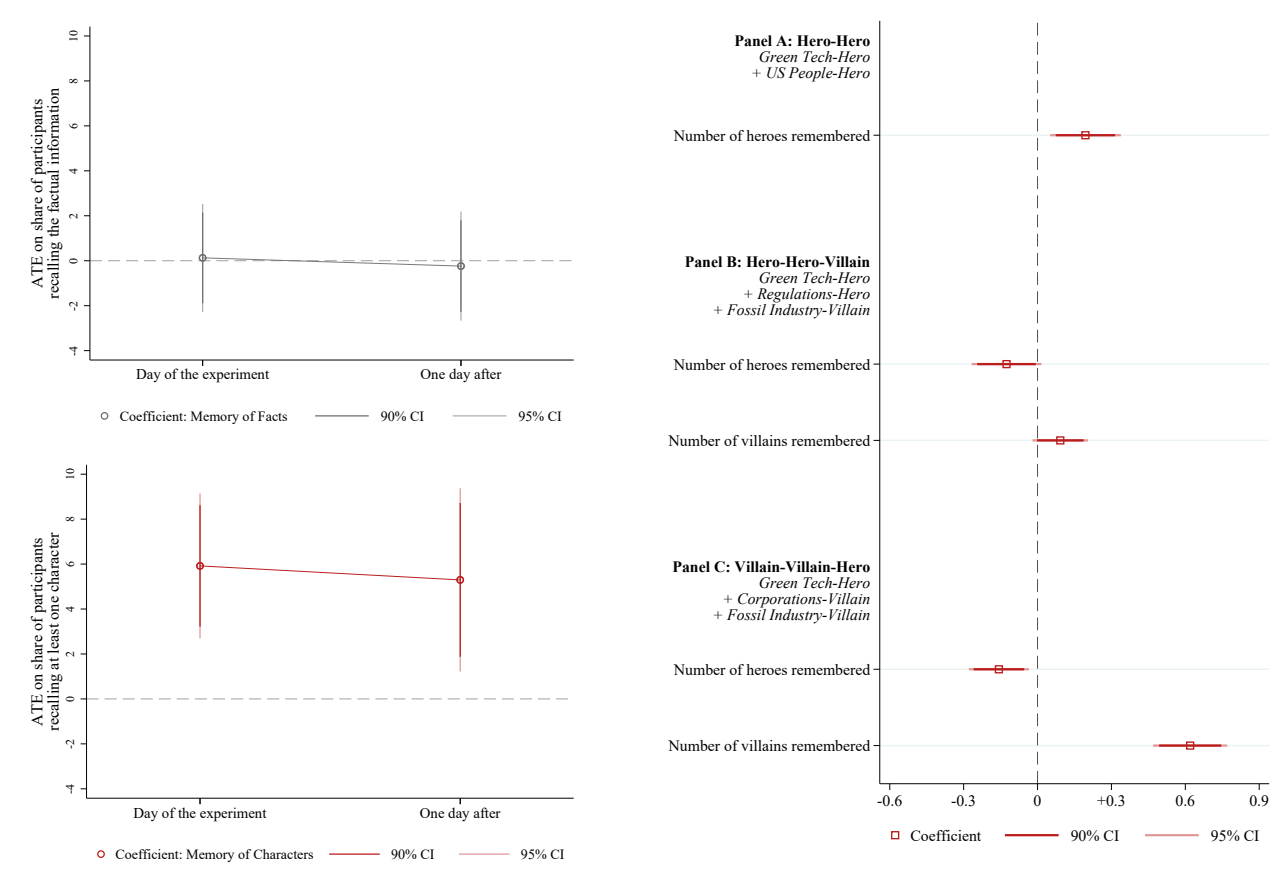


Notes: The figure reproduces the results shown in the paper Figure 13 but excluding controls.

Figure L.4: ***Experimental Results-The Impact of Political Narratives on Memory (Pooled Sample, No Controls))***

(a) Memory: Recall of Factual Information and Character Presence

(b) Memory: Recall of Characters by Role



Notes: The figures reproduce results in the paper exhibit Figure 14 but excluding controls from all regression models.

L.6 The Impact of Valence and Anger

One potential concern with our *persuasion 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.

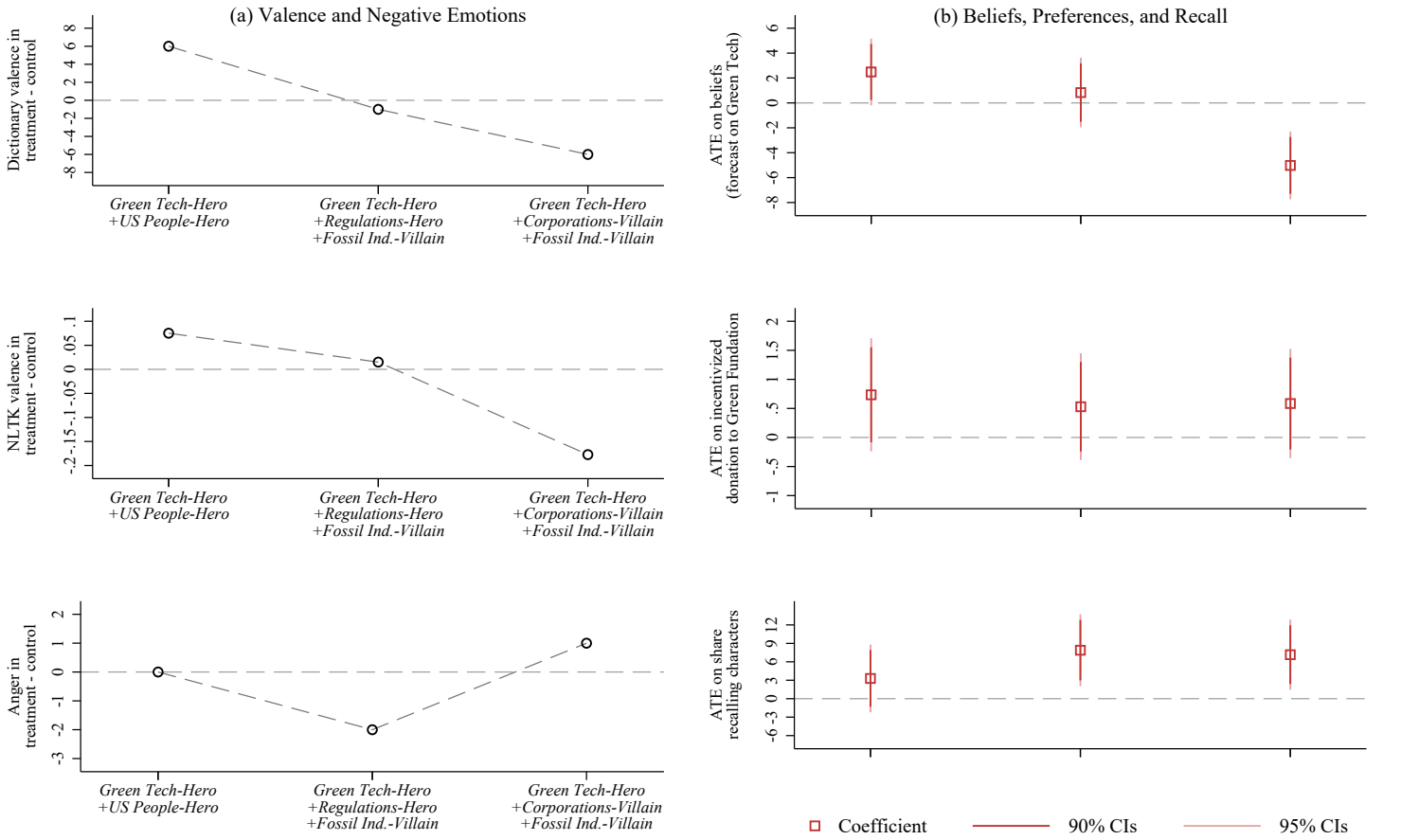
While we recognize that emotions are an important ingredient in building and communicating political narratives, our analysis throughout the paper shows that they are not sufficient to explain their effect. The roles of villain, hero, and victim capture deeper mental constructs that extend beyond emotional 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 L.5 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 L.5 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 L.5: ***Emotional Content in the Experimental Design***



Notes: Panel a of Figure L.5 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 L.5 panel (b) replicates the main results of the paper. Reference figures are Figure 12a, Figure 13, Figure 14b.

L.7 Randomization Inference: Persuasion Experiments

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 L.4: ***Experimental Results-The Impact of Political Narratives on Beliefs-Randomized Inference***

Dependent Variable	**Expectation for the Future**		
	Hero-Hero	**Hero-Hero-Villan**	**Villain-Villain-Hero**
	Coeff./SE/*Randomized (n=1000)* p-value		
	(1)	(2)	(3)
Treatment	2.487	0.831	-5.022
	(1.368)	(1.425)	(1.387)
	[0.077]	[0.546]	[0.000]
Controls	✓	✓	✓
Experiment: Hero-Hero	✓		
Experiment: Hero-Hero-Villain		✓	
Experiment: Villain-Villain-Hero			✓
Mean Outcome Control Group	39.02	43.67	42.32
Observations	987	976	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 12a in the paper is the reference plot.

Table L.5: ***Experimental Results-The Impact of Political Narratives on Stated Preferences-Randomized Inference***

Dependent Variable	**Stated Preferences (Policy Support)**		
	Hero-Hero	**Hero-Hero-Villan**	**Villain-Villain-Hero**
	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 12b in the paper is the reference plot.

Table L.6: ***Experimental Results-Impact of Political Narratives on Revealed Preferences-Randomized Inference***

Dependent Variable	**Revealed Preference: Incentivized Donation**			
	Hero-Hero	**Hero-Hero-Villan**	**Villain-Villain-Hero**	**Pooled Sample**
	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 13 in the paper is the reference plot.

Table L.7: ***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 14a are the reference plots.

Table L.8: ***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 14b in the paper is the reference plot.

M Memory Channel: Interpretation and Discussion

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 L.1), 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, C. 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, 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, C. 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, C. 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 L.5 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. If most recipients of narratives only

remember the characters and their roles better, but not the facts, this could lower the perceived costs of spreading imprecise 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.