Gehring, Kai; Grigoletto, Matteo

Working Paper

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

CESifo Working Paper, No. 12064

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Gehring, Kai; Grigoletto, Matteo (2025) : Virality: What Makes Narratives Go Viral, and Does it Matter, CESifo Working Paper, No. 12064, Munich Society for the Promotion of Economic Research - CESifo GmbH, Munich

This Version is available at:
https://hdl.handle.net/10419/327674

CES ifo
Working Papers

12064
2025

August 2025

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

Kai Gehring, Matteo Grigoletto

CES ifo

Imprint:

CESifo Working Papers

ISSN 2364-1428 (digital)

Publisher and distributor: Munich Society for the Promotion
of Economic Research - CESifo GmbH

Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740

Email office@cesifo.de
https://www.cesifo.org

Editor: Clemens Fuest

An electronic version of the paper may be downloaded free of charge
· from the CESifo website: www.ifo.de/en/cesifo/publications/cesifo-working-papers
· from the SSRN website: www.ssrn.com/index.cfm/en/cesifo/
· from the RePEc website: https://ideas.repec.org/s/ces/ceswps.html

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

Kai Gehring† and Matteo Grigoletto‡

August 12, 2025

Abstract

Understanding behavioral aspects of collective decision-making remains a core challenge in economics. Political narratives can be seen as a key communication technology that shapes and affects human decisions beyond pure information transmission. The effectiveness of narratives can be driven as much by their virality as by their specific persuasion power. To analyze political narratives empirically, we introduce the political narrative framework and a pipeline for its measurement using large language models (LLMs). The core idea is that the essence of a narrative can be captured by its characters, which take on one of three archetypal roles: hero, villain, and victim. To study what makes narratives go viral we focus on the topic of climate change policy and analyze data from the social media platform Twitter over the 2010–2021 period, using retweets as a natural measure of virality. We find that political narratives are consistently more viral than neutral messages, irrespective of time or author characteristics and other text features. Different role depictions differ in terms of emotional language, but political narratives capture more than merely valence or emotions. Hero roles and human characters increase virality, but the biggest virality boost stems from using villain roles and from combining other roles with villain characters. We then examine the persuasiveness of political narratives using a set of online experiments. The results show that narrative exposure influences beliefs and revealed preferences about a character, but a single exposure is not sufficient to move support for specific policies. Political narratives lead to consistently higher memory of the narrative characters, while memory of objective facts is not improved.

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

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

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

†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 complexity about human or instrument characters into three archetypal roles called the drama triangle (hero, villain, victim) that can be processed, remembered and, crucially, shared easily. Hence, political narratives have an efficiency advantage in information markets when attention is limited and information processing costly. This article investigates (i.) which features enhance the virality of a political narrative and (ii.) the impact of political narratives on beliefs, preferences and memory.

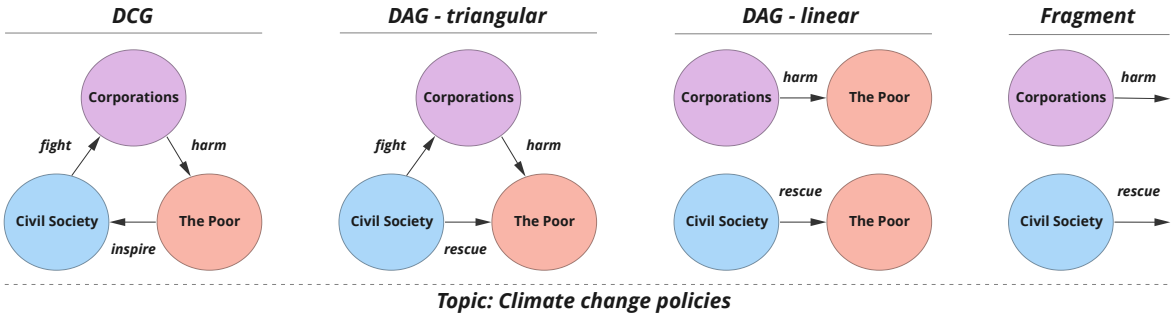
In popular science books (Harari 2014; Shiller 2020a), but also increasingly in economics (e.g., Bénabou, Falk, and Tirole 2020; Bursztyn et al. 2023b; Esposito et al. 2023) narratives are now generally recognized as a key communication technology that influences human preferences, beliefs, and decisions.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

This enables us to study the two key features of narratives: virality and persuasive power. We study virality where it is most visible: the retweet metrics on the social media platform Twitter/X. We analyze 1.15 million climate change policy tweets over the 2010 to 2021 period, encoding twenty 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 conditional on author characteristics, text quality and emotions. Villain and hero roles produce the largest individual boost and combinations that add extra villain characters further raise virality. In complementary, pre-registered survey experiments with 3000 respondents, political narrative exposure shifts beliefs and real-money 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 both from repeated exposure to viral narratives, as well as from individual persuasiveness and improved memory.

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

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

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



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

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

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

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

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

To explore our framework's capabilities, we pre-define ten characters, five *human* (e.g., corporations, U.S. people) and five *instrument* (e.g., fossil industry, green technology, regulation) 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) to collect over three million English-language tweets via the Twitter API, sampling every Saturday plus one randomly chosen day per month from 2010–2021. We further use the metadata to select tweets originating from the US, ending with roughly 1.15 million tweets. Then we apply our pipeline, using the GPT-4o-mini model from OpenAI, to first further validate whether a tweet fits the topic. For each in-topic tweet, we detect whether each pre-defined character is present and, if so, whether it appears neutrally or in a drama triangle role: hero, villain, or victim.

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

We begin with descriptive results on political narrative frequency over time, which show marked

shifts between 2020 and 2021. The Green Tech–Hero character-role is most frequent, followed by Fossil Industry–Villain and Corporations–Villain, while the U.S. public appears often as both Hero and Victim, with notable changes over time. Beyond individual character–roles, the co-occurrence within a tweet of multiple roles reveals details about the political narratives over that time period. Taken together, these patterns provide an initial map of how political narratives - measured through the character–roles and their combinations – evolve in the climate-policy discourse.

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

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

Our observational analysis has several limitations. First, Twitter/X was widely used in the U.S. across partisan lines during our sample period, but users do not systematically represent the general U.S. population. Second, the climate change policy discussion in the U.S. 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 tractable first setting; the measurement framework and pipeline are general and can be applied to other topics and corpora. Third, our analysis of virality

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

is correlational: despite extensive controls and fixed effects, selection and confounding cannot be entirely ruled out. Fourth, platform ranking and the Twitter algorithm can itself shape what is seen and shared. Controlling for sentiment, emotions, and language features somehow limits the issue, but algorithmic amplification remains a residual concern that cannot be avoided when studying social networks. Fifth, automated accounts (bots) may affect diffusion dynamics; we apply standard filters for robustness tests, but acknowledge that social bots remain an endemic feature of social media.

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

Contributions: This paper contributes to various strands of literature. First, we contribute to the growing "narrative economics" literature (e.g. Shiller 2017; Bénabou, Falk, and Tirole 2020; Eliaz and Spiegler 2020) that emphasizes the critical role of narratives in shaping economic phenomena and decision-making processes. Despite the increasing interest, there is currently no coherent and systematic definition or measurement of narratives in the context of economic research. To address this gap, we propose a framework to define and measure political narratives. Drawing on insights from various disciplines ranging from political science (Terry 1997; Verweij et al. 2006; Jones and McBeth 2010; Jiangli 2020), sociology (Brittain 2006; Polletta et al. 2011; Merry 2016; O'Brien 2018), communication studies (Anker 2005; Gomez-Zara, Boon, and Birnbaum 2018) to literature (Fog et al. 2010), our framework provides a clear and specific approach to understanding and measuring narratives. Unlike narrower definitions that focus solely on causal or temporal chains, our approach encompasses a wider range of narratives. Additionally, unlike broader definitions that may lack specificity, our framework provides a clear and systematic way to identify and classify narratives. This paper offers an efficient approach for researchers to more effectively understand and measure narratives in any type of text data.

Our political narratives framework, which incorporates insights from various disciplines can help to bridge the gap between narrative analysis approaches in diverse economic fields. Recent empirical research investigates the impact of narratives across diverse topics, advancing our understanding of their role in political economy and macroeconomic contexts. This body of work spans subjects such as racism (Esposito et al. 2023), diverging narratives during the COVID-pandemic (Bursztyn

et al. 2023b), and monetary policy perception and inflation (Andre et al. 2025; Macaulay and Song 2022). Recent studies in political economy analyze "folklore" (Michalopoulos and Xue 2021) and movies (Michalopoulos and Rauh 2024). Flynn and Sastry (2024) provide evidence that contagiously spreading narratives drive a significant share of the business cycle.[2] Overall, there is an increasing interest in macroeconomics and finance to understand how collective behavioral phenomena like animal spirits (Akerlof and Shiller 2010) can drive economic decisions. Our framework can be adapted to study any type of narrative that has the purpose of changing preferences and beliefs. For instance, central banks could be framed as heroes or as villains in the fight against inflation, or, not rare in contemporaneous politics, they might be the victims of intrusive governments.

Third, by examining the frequency and virality of political narratives about climate change policy in real social media interactions, our results provide many additional insights and more nuances to an emerging and quickly growing literature on the political economy of climate change. Dechezleprêtre et al. (2025) demonstrate that, in a global survey of 20 countries, support for climate change policies hinges more on an individual's perception of the policy's effectiveness on their well-being and political orientation, rather than on their knowledge or understanding of climate change facts. Andre et al. (2024) highlight the role of social norms, moral values, and the perceived behavior of others in determining the willingness to financially support climate change policy. Our experimental results shows the effects of even single exposure to a political narrative, which is then further amplified by the increased virality of political narratives.

Our results on virality and persuasive power also contribute and link to important strands of literature in other disciplines that we cannot fully acknowledge here. Berger and Milkman (2012) is one of many studies that investigate virality. They show that New York Times articles which lead to high arousal and activate emotions tend to be more viral. Compared to merely studying emotions or other specific text features, our framework allows us to better understand structural and systematic features driving virality, based on the drama triangle as a millennia-old archetypal classification scheme. In political science, Bernauer (2013) examine communication strategies for garnering public support for climate change, and Huber, Greussing, and Eberl (2022) use a conjoint experiment and a survey to study public support for climate change policies. Bernauer (2013) provides an overview of climate change research in political science.

Second, we introduce and evaluate an efficient methodology and pipeline to implement the political narrative framework in empirical applications. Our approach builds on the foundational contributions of Gentzkow, Kelly, and Taddy (2019) and O'Neill et al. (2021), who provide comprehensive overviews of text-as-data methods for economists, as well as on narrative analysis tools developed by Ash, Gauthier, and Widmer (2024). Specifically, we leverage LLMs via a structured pipeline and batch-level API calls to identify characters and their roles within large amounts of text. We show that using the political narrative framework captures important nuances beyond emotions (Gennaro and Ash 2022) and other text features that have been studied before. In line

[2] While not specifically framing it as a narrative, Cagé et al. (2023) relate to the concept of heroes and villains by studying the impact of French general Petain's perception as a combat hero on political views and behavior during WWII.

with other recent paper employing LLMs in innovative ways (Lagakos, Michalopoulos, and Voth 2025; Voth and Yanagizawa-Drott 2025; Caesmann et al. 2024), we demonstrate how LLMs allow us to move beyond simple metrics like emotions and study more structural features systematically. Our approach is applicable to basically any topic and can, in contrast to prior, mostly manual, approaches for purely causal narratives, be scaled to large text corpora.

Finally, our work contributes to a large literature in media economics (see overview in Zhuravskaya, Petrova, and Enikolopov 2020) and the economics of social media (see overview in Aridor et al. 2024), by providing a novel framework and pipeline to analyze narratives in the media. While initial seminal studies focus on newspapers (Gentzkow and Shapiro 2010) more recent work investigates TV (Ash et al. 2024a; Ash et al. 2024b; Ash and Galletta 2023; Ash and Poyker 2024), the internet more broadly (Cagé, Hervé, and Viaud 2020; Campante, Durante, and Sobbrio 2018) and social media specifically (Acemoglu, Hassan, and Tahoun 2018; Enikolopov, Makarin, and Petrova 2020; Müller and Schwarz 2023). In contrast to previous findings that emphasize the limited persuasive power of facts compared to narratives (Bursztyn et al. 2023a) and their inability to change policy conclusions from right-wing fake news (Barrera et al. 2020), our approach facilitates a more nuanced exploration of narrative dynamics. By building on examples such as Campante, Durante, and Sobbrio (2018), which demonstrates how politicians construct narratives by connecting unrelated events, our work paves the way for future research to uncover the underlying mechanisms driving the impact of narratives on public opinion and policy discourse.

2 Defining and Measuring Political Narratives

2.1 Prior Approaches

Since economists began studying narratives, they have successfully developed formal theoretical models and applied quasi-experimental approaches to establish causality in specific contexts. Initially, definitions of narratives in the literature were broad and lacked a common understanding. Early contributions, such as Shiller (2020); Shiller (2017), defined narratives loosely. For instance, Shiller (2017) describes a narrative as "[. . .] a simple story or easily expressed explanation of events that many people want to bring up in conversation" Shiller (2017), p. 968. In this work, virality was also considered a defining feature of a narrative. Many empirical papers adopt similarly broad definitions. For example, Esposito et al. (2023) describe narratives as the "framing of memories and selective recollection of facts[. . .]" (p.2), while Bursztyn et al. (2023) use the concept without imposing a restrictive definition.

While broad definitions provide flexibility, they also create challenges. The lack of a common framework makes it difficult to compare empirical results and limits systematic measurement. Without a clear structure, communication becomes unnecessarily complex, and replication across studies remains difficult. To address these issues, economists have moved toward narrower, more structured definitions that facilitate empirical analysis and improve reproducibility. One widely adopted approach defines narratives as temporal or causal sequences.

Modeling narratives as directed graphs of sequences between different entities (**andre˙narratives˙2024**) or as mappings of "actions to consequences" (Eliaz and Spiegler 2020, p.2) marks an important step forward. As shown in our example in the introduction, directed acyclic graphs (DAGs) help structure and visualize narratives by ordering variables according to an underlying causal or temporal model. Similarly, Ash, Gauthier, and Widmer (2024) develop a package that extracts all possible subject-verb-object sequences from selected texts, capturing a subset of narratives that fit a specific DAG structure.

While this approach provides analytical clarity, it is also restrictive, as many widely shared narratives do not conform to a strictly causal or temporal structure. In social media, political discourse, and news coverage, narratives often lack explicit causal links but remain powerful in shaping perceptions. Statements like "corporations are greedy" or "immigrants cannot be trusted" do not follow a clear sequence, yet they function as influential and widely shared narratives. These examples illustrate the limitations of purely causal approaches and underscore the need for a framework that accounts for narratives beyond those strictly defined by causal links.

2.2 Political Narratives

In this paper, we introduce a definition of narrative that complements the existing literature in economics. In particular, we define and analyze what we call political narratives, narratives framed within a specific topic and built around their characters featured in a particular role. Our definition is complementary to the commonly used definition of narratives as causal sequences, connecting events and characters.

This definition builds on a long tradition of research across disciplines. The drama triangle (Karpman 1968) is widely recognized as a central character set in human storytelling. While other character schemes exist, most additional roles remain peripheral, gaining relevance only in relation to these three core roles. Moreover, while social identities and roles evolve over time (Bergstrand and Jasper 2018), the drama triangle has remained remarkably stable. The archetypal hero fighting the villain to save the victim appears across cultures and historical periods. Empirical work by Bergstrand and Jasper (2018) further supports this framework, showing that most character traits can be clustered within these three roles, aligning with how humans intuitively comprehend, interpret, and recall stories.

The drama triangle hero-villain-victim goes beyond simply capturing sentiment in text. Nevertheless, emotionally charged language is often an integral part of these kinds of narratives. While the drama triangle simplifies narrative structure, focusing on these three archetypal roles, it often lends itself to narratives where there is a clear moral of the story. Political narratives within the drama triangle framework often resemble slogans or sensationalistic content. But they can also provide a clear consequential structure to the narrative.

Table 1 provides insights into the differences and similarities between narratives defined as causal sequences and political narratives. Many features may identify and define narratives, but in the first three columns of the table we report three core ones: sequences, emotions, and the presence

Table 1: *Examples of Narratives*

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

of character-roles. We provide both general and climate related examples and we show how some can be considered only causal narratives, other only political, some none of the two and some both.

2.3 Measuring Narratives

Text analysis methods used in economics have traditionally fallen into two broad categories: dictionary-based approaches and machine learning methods. Dictionary methods are straightforward, analyzing text by counting the frequency or share of predefined keywords. This approach can be particularly effective when pre-defined dictionaries exist for capturing the sentiment of specific emotions. However, the simple word-counting mechanism lacks the capacity to account for context and complex sentence structures. To address these limitations, more recent approaches integrate Natural Language Processing (NLP) tools with dictionary methods, allowing for a more nuanced understanding of word usage, sentence dependencies, and textual relationships (Gehring, Adema, and Poutvaara 2022).

By contrast, unsupervised machine learning methods, such as topic models, take a more flexible approach by clustering similar text elements. These models generate latent topics based on patterns in the data, but users must define parameters and interpret the results manually. After running the model, researchers must distill, label, and validate the topics, often requiring subjective judgment. More advanced tools, such as RELATIO (Ash, Gauthier, and Widmer 2024), are very valuable for exploration, but still require significant human input. Linking back to the DAGs we depicted before, identifying causal arrows relies on first defining the nodes linked by the arrows. Using a package like RELATIO, requires users to eventually condense and aggregate a large set of entities, the nodes,

into a smaller, more relevant and manageable subset. This underscores a fundamental limitation of any automated method: regardless of automation, human interpretation remains crucial in text analysis. In our approach, this dimension-reduction happens at the beginning, requiring user to justify their character choices and thus ensuring transparency.

Recent breakthroughs in Natural Language Processing (NLP), particularly the rise of Large Language Models (LLMs), have transformed text analysis. These Transformer-based models, such as OpenAI's GPT4O-mini, have dramatically improved the ability to process, interpret, and generate human-like text. LLMs have already enhanced productivity across various domains and they are increasingly valuable for academic research, including the type of analysis conducted in this paper. Unlike traditional methods, LLMs provide both speed and scale, making them particularly suited for large-scale narrative analysis. The accessibility of APIs allows researchers to interact with these models dynamically, much like a chatbot, but with the added ability to process massive amounts of text data at unprecedented speed. This combination of interactive flexibility and computational power marks a fundamental shift in the way researchers analyze narratives, thus making this method our preferred choice.

2.4 Pipeline

This section provides a brief summary of the pipeline we followed for the creation of the final dataset used in the analysis. The pipeline comprehends five steps: (1.) selecting and defining the topic of interest, (2.) identifying a source and extracting data, (3.) identifying relevant characters, (4.) preparing a prompt (5.) obtaining predictions.

1. Selecting and defining the topic of interest: A well-defined topic is a prerequisite for a fruitful narrative analysis. The clearer the topic, the more straightforward the identification of relevant characters and the exploration of the research question. In our application, we analyze the political economy of climate change. A review of the relevant literature reveals two dominant discussions: scientific evidence on climate change and policy responses. Since our focus lies within the domain of political economy, we concentrate on climate change policies, deliberately excluding debates on the scientific reality and predictability of climate change.

2. Identifying data sources and extracting data: After selecting the topic, the next step consists in gathering data. The most common sources of text data in economics are digitized newspapers and social media. However, researchers can also consider resources like transcribed TV, radio, and YouTube broadcasts, or open-ended responses from surveys for more unique insights. For this study, our focus is on narratives about climate change policies in the United States, collected from the social media platform Twitter/X over the period 2010-2021. We choose the U.S. due to the significant role Twitter plays in shaping and disseminating narratives there. The data collection process involves querying the Twitter historical APIv2 with a set of keywords adapted from Oehl, Schaffer, and Bernauer (2017).

3. Identifying characters: The third step represents a critical stage in the pipeline, centered on identifying relevant characters within the topic. Character selection is primarily guided by the

research question and the researcher's analytical focus. However, it is crucial to balance the scope of selection with practical considerations. Expanding the set of characters too broadly can complicate prediction accuracy and increase computational demands, while a more focused selection ensures greater reliability and interpretability of results.

Characters can be identified through various methods, including literature review, topic modeling, and entity recognition tools. In our case a mix of these methods lead us to define a list of five 'human' characters – made of institutions and groups of individuals – and five 'instrument' characters – meaning policy tools and instruments. The human characters are: Developing Economies, U.S. Democrats, U.S. Republicans, Corporations, and the U.S. People. The instrument characters are: Emission Pricing, Regulations, Fossil Industry, Green Tech, and Nuclear Tech.

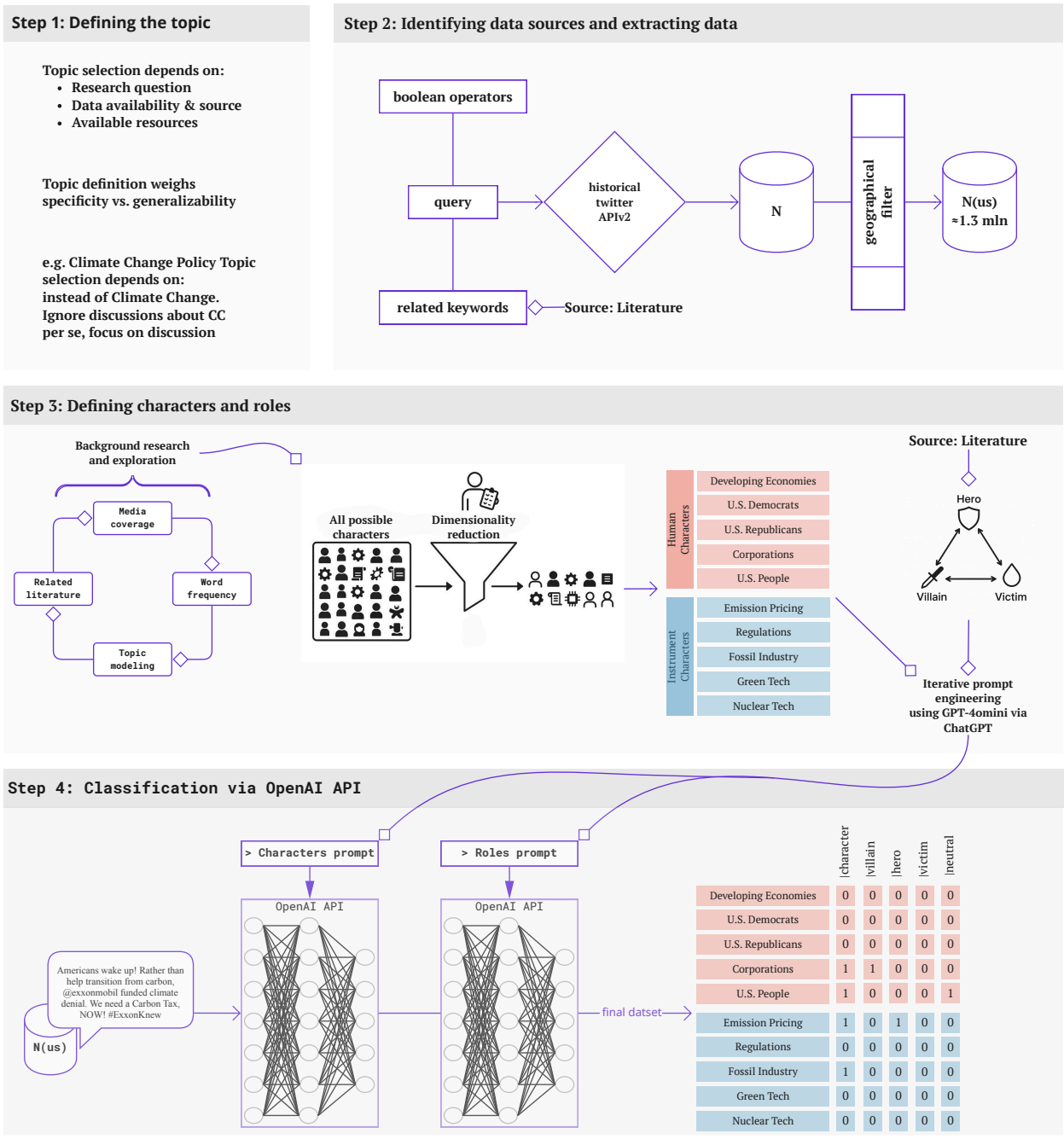
4. Preparing a prompt: The fourth step involves designing an effective prompt to instruct the AI model. Large Language Models (LLMs) like GPT allow researchers to send thousands of queries through API calls using a single standardized prompt. Prompt engineering is a crucial part of the pipeline, as the quality of the prompt directly affects the accuracy and consistency of the model's responses.

We recommend designing the prompt with the assistance of the same model that will later annotate the text data. In our case, we used GPT-4o via OpenAI's ChatGPT to iteratively refine the prompt. We first provided the model with a description of the task's objective and then optimized the prompt through multiple iterations to ensure clarity and precision. The final prompt(s) used in this study are provided in the Appendix.

5. Obtaining the predictions: The final step of our pipeline consists of obtaining predictions for each tweet, as visualized in Figure 1. We use the OpenAI API to process every tweet in our dataset with GPT-4o-mini, applying the same prompts consistently across all observations. To improve efficiency, we parallelize the querying process, allowing thousands of tweets to be annotated simultaneously. The entire process cost approximately $200 and took two days of continuous computation, making it an extremely cost-effective approach for annotating over one million tweets.

Conceptually, the annotation process can be understood as the OpenAI API acting as a classification engine, taking in specific inputs and generating structured outputs. Each tweet is processed in its original form, along with the definitions of the drama triangle roles —- hero, villain, and victim -— as well as the list of characters identified in Step 3 and the prompt designed in Step 4. Based on this input, the model predicts first the presence of the characters. After, it assigns roles to the relevant characters in each tweet. If a character is explicitly present, the model determines whether it fits the role of hero, villain, or victim. If the character appears in the text but does not conform to any of these roles, it is instead labeled as neutral. The output is a structured dataset in which each tweet is annotated according to this role-based classification, allowing for a systematic and replicable narrative analysis.

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



Notes: The figure shows a schematic representation of the pipeline used for the annotation of tweets. The process is applied to each tweet, and it is parallelized to run on multiple thousands tweets simultaneously.

3 Data: Twitter Sample over 2010-2021 Period

The first step of our data collection consists in extracting tweets containing climate change related keywords, that we adopt from Oehl, Schaffer, and Bernauer (2017). We collect tweets from each Saturday between 2010 and 2021, in addition to an additional data extraction on one randomly

selected day every month (details in Appendix A). The goal is to limit the amount of tweets to a number that makes analysis feasible, while making the dataset as representative as possible. We exclude non-English tweets and retweets while querying the Twitter API, resulting in a set of exclusively original tweets in English. We conduct ex-post checks to drop potential duplicates, decreasing the number to a total of 3,279,730 non-duplicated tweets.

Our analysis concentrates on discussions about climate change policies within the United States. Accordingly, we include only tweets originating in the U.S., resulting in a final dataset of 1,151,671 tweets. We select the U.S. as our focus due to its status as the world's second-largest emitter of CO_2, the ready availability of social media data, and vibrant public debate on climate-change policies. The U.S. discourse on climate change is particularly contentious, with Twitter serving as a key platform of debate over the observed time period. Because tweets typically lack inherent geographical information, we employed two methods to determine their location: first, by leveraging location details provided in user profiles; and second, by using geo-localization tags—available on a minority of tweets, that allow users to "tag" their posts with a specific coordinate box. Appendix A.2 provides more details.

C.3 provides descriptive statistics for the analysis of tweets originating from the U.S.; the average number of words in the tweets of our sample is 23, excluding hashtags and mentions. The average tweet in our sample is liked 11 times and re-tweeted 2.2 times. There is a lot of variation in the sample, suggesting that these differences are an important feature of a tweet. A standard deviation of 1229 shows the strong skewness of the data.

We present the results of the annotation done via GPT-4o in 2. The table shows the shares of character-roles that are identified in the tweets for all character-role combinations. Panel (a) presents the results for characters that we define as human, like U.S. People, while panel (b) shows the results for instrument characters. The most prevalent instrument character-roles are Green Tech portrayed as a hero (11.51% of all tweets) and the Fossil Industry in the role of a villain (11.75%). The most commonly identified characters for the human characters are the neutral representation of U.S. People (about 9.7 % of all tweets), the U.S. Republicans as villains (9.5%), and Corporations as villains.[3]

[3] The extent to which Twitter data is representative of broader public discourse remains unclear. To shed light on the external validity and prevalence of character roles in other media debates, we conducted a validation exercise. We sourced the 1,000 most prominent articles from the three most widely circulated U.S. daily newspapers. The articles were divided into three-sentence snippets, and each excerpt was classified using the same prompting scheme as the Twitter data. The results of this exercise are presented in Appendix H. Remarkably, we find that the shares and distributions of character roles are highly representative of the Twitter debate. For instance, the three most common character roles are the green-tech hero, Corporations-villains, and the Fossil Industry-villain with similar percentage shares as the main tables.

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

Panel A: Human Characters

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

Panel B: Instrument Characters

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

Notes: The table displays the frequency of character-roles in the classified data. Shares are computed considering only 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 in the computation of shares only character roles that appear at least 100 times, thus excluding U.S. 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 tweet but is not depicted in one of the three specific roles. The occurrence of character-roles is not mutually exclusive, meaning multiple roles may appear in the same tweet. The appendix Table C.6 shows the frequency in absolute numbers, including also the categories excluded here, because not reaching the 100 instances in the time period.

4 Descriptive Evidence

4.1 Features of Political Narratives

We begin by examining the text features of the political narratives, featuring a topic and characters as neutral or in a hero, villain, or victim role. The way these narratives are constructed can vary significantly. Many elements such as emotional expression, causal links among characters, the plot of the story, the tone, and the quality of writing can influence how narratives are perceived. We focus here on two key aspects, emotional language and text quality, and highlight how they relate to the different roles that are at the core of our framework.

The goal of this analysis is to examine whether tweets containing narratives differ from those where characters appear only in a neutral role. To do so, we analyze all tweets mentioning at least one character and compare their emotional content, valence, and textual characteristics. Specifically, we distinguish between tweets where characters are framed as heroes, villains, or victims and those where characters are mentioned without a narrative role. Appendix Figure C.2 provides an overview of the composition of the data set. Approximately 16% of the climate policy tweets do not mention any identified characters, while 15% feature only neutral characters. The remaining tweets contain at least one character assigned to a narrative role. In this and all subsequent analyses, we focus on

tweets with at least one character, so that we can really focus on comparing different text features of characters in a neutral compared to the other roles.

Figure 2 provides an overview of emotional expression and text quality in narrative vs. neutral tweets. The top panel presents the average scores for two key aspects - emotions and valence on the left, and text quality on the right. To measure emotions we follow established dimensions and measure the frequency of words associated with joy, trust, anticipation, surprise, anger, disgust, fear, and sadness. Valence refers to the general positivity or negativity of a text, which we also approximate with an established dictionary. All dictionaries are taken from the widely cited and validated NRC Dictionary (Mohammad and Turney 2013).

Regarding text quality metrics, we measure several linguistic features including reading ease, expected education level required for comprehension, type-token ratio – a measure of language diversity – number of words per tweet, average word length, and average sentence length. The bottom panel presents the same information using a spider chart, breaking down narrative tweets by role. This visualization compares hero, villain, victim, and neutral roles, where points closer to the center represent lower scores, while those further away indicate higher values. All scores are standardized to facilitate comparison and improve readability.

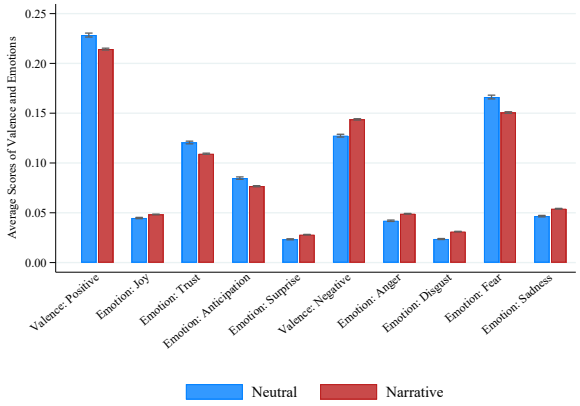
Some noticeable characteristics of narratives stand out. When comparing narratives and neutral tweets it is evident that neutral tweets contain more positive language, both in terms of general valence and individual NRC emotions. In contrast, narratives tend to exhibit more negative language, both in overall valence and specific emotions, with the exception of fear. This somewhat puzzling pattern becomes clearer in the spider chart. Breaking narratives down by role reveals substantial variation in positive valence and emotions across different character roles. Hero narratives are the most positive, but the overall narrative average is lowered by the lower scores of victim narratives. Victim narratives also contain the highest levels of fear, while villain narratives are the most negative. In general, narratives tend to carry a stronger emotional charge compared to neutral tweets, with the exception of trust and anticipation, which remain more prevalent in neutral content.

A different pattern emerges when examining text quality. Narratives and neutral tweets show no substantial differences across nearly all dimensions. The only notable exception, when comparing narratives to neutral tweets overall, is that narrative tweets tend to be longer on average. Even when breaking down by role, differences remain minimal, except in sentence length. Victim narratives consistently appear as the longest tweets, while hero narratives tend to be the shortest.

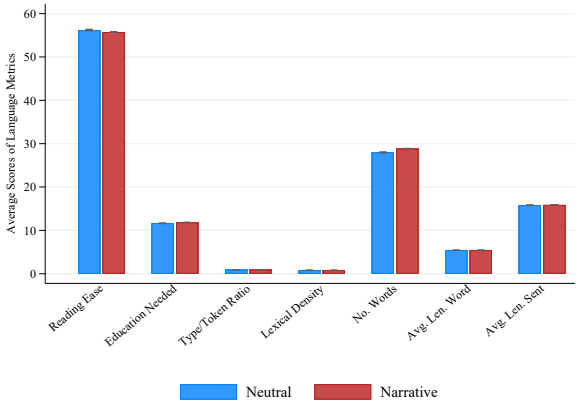
Overall, this analysis suggests that narratives differ from neutral tweets in their emotional intensity and linguistic characteristics. Narrative tweets tend to exhibit stronger emotional variation across roles, with hero narratives being the most positive and villain narratives the most negative. In terms of text quality, differences remain minimal, apart from narrative tweets being slightly longer on average, particularly for tweets containing victim narratives.

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

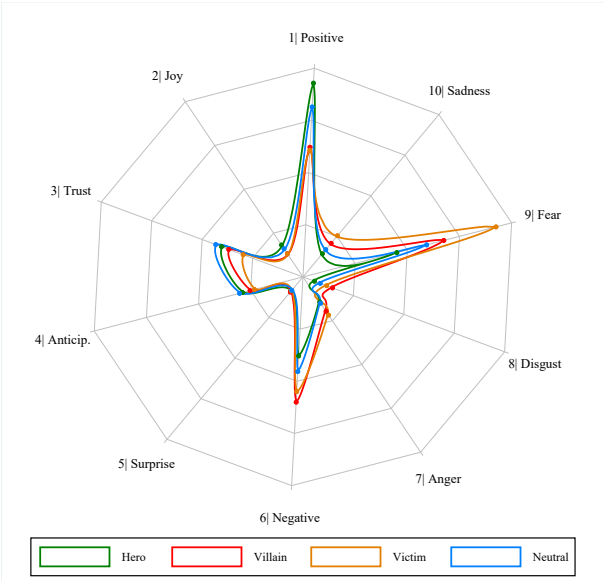
(a) Neutral vs. Narrative: Emotions and Valence



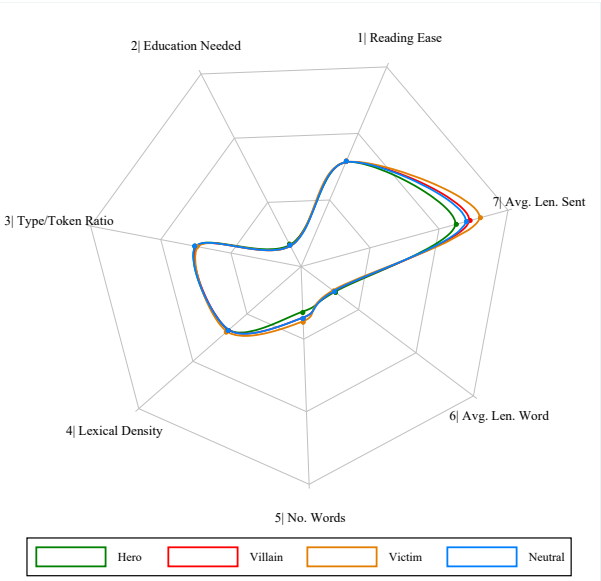
(b) Neutral vs. Narrative: Language Metrics



(c) Comparison Across Roles: Emotions and Valence



(d) Comparison Across Roles: Language Metrics



Notes: Figure 2a shows the average levels of valence and emotions in tweets containing at least one character-role compared to tweets containing characters in a neutral role. Figure 2b presents the same comparison for several measures of language metrics. Figures 2c and 2d show the average of the measures in radar charts, distinguishing neutral from each role in the drama triangle: hero, villain, and victim. In each radar chart, smaller values are positioned toward the center, while larger values extend outward, and the measures are standardized to ease readability. In all figures, valence is computed as the average amount of positive and negative NRC words. Emotions represent the average occurrence of words from the NRC Dictionary, specifically joy, trust, anticipation, surprise, anger, disgust, fear, and sadness. Language metrics include reading ease – which indicates the simplicity of the text – education needed for comprehension – estimated as the years of education required to understand the text – type/token ratio – which measures word diversity – lexical density – which reflects the proportion of content words to total words – number of words per tweet, average word length, and average sentence length.

4.2 Frequency of Political Narratives

After showing the inherent difference between narratives and neutral tweets, we now move to explore the frequency with which different character-roles appear and how their distribution evolves over time. To examine this, we apply our pipeline to a dataset of more than one million English-language tweets related to climate policy in the United States. The dataset spans January 2010 to September 2021, a period that includes significant economic and political shifts, as well as key climate negotiations and policy decisions.

Figure 3 presents the share and evolution of climate policy narratives over time. The figure captures the frequency of individual character-roles appearing in tweets but does not yet account for their co-occurrence within the same tweet, which we explore below. As of 2021, the most common character-roles on average are Green Tech-hero (23%) and Fossil Industry-villain (16.9%), followed by U.S. Republicans-villain (11.8%), Corporations-villain (10.7%), U.S. Democrats-hero (6.7%), U.S. People-hero (5.2%), and Regulations-hero (4.6%), with all other roles appearing less frequently.

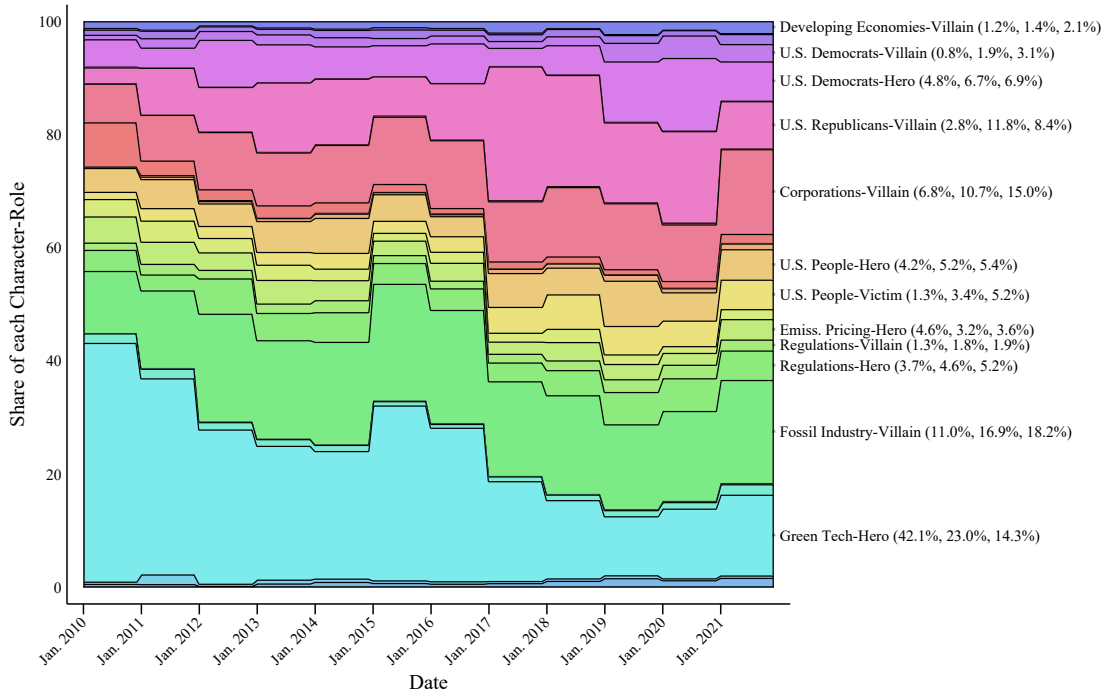
The composition of narratives shifts considerably over time, reflecting broader changes in climate policy discourse. Fossil Industry as a villain became increasingly prevalent, rising from 11% to 18%, while U.S. People as victims, though in general not very common, nearly tripled in frequency, suggesting growing public discontent. A parallel trend is observed in the portrayal of Corporations as villains, which increased from 7% to 15% by 2021, reinforcing the perception of corporate actors as central contributors to climate-related challenges. One of the most striking shifts concerns the decline of the Green Tech-hero narrative, which fell from 42% to 14%, potentially signaling a decreased optimism about technological solutions to climate change. At the same time, the increase in U.S. Republicans framed as villains suggests a more politicized debate in recent years, reflecting heightened polarization in climate policy discussions.

Taken as a whole, these changes suggest a growing frustration with government inaction or even perceived support for fossil industries and other contributors to climate change.[4] Beyond a declining reliance on the government as a hero, we observe a marked increase in the Corporations-villain narrative, coupled with decreasing trust in Green Tech-hero. Overall, the discussion appears to have become more politicized and polarized, reflecting a gloomier outlook on the future.

While Figure 3 shows the overall frequency with which a specific character-role appears, Figure 4 depicts possible combinations of that with other character-roles, however limited to tweets with a maximum number of two character-roles, to ease the visualization. The figure shows all combinations in a matrix form, with the diagonal reporting occurrences of the character-role in simple narratives, hence without any other character-roles. For the remaining entries of the matrix, we compute for each single character-role the amount of times it appears in a 1- or 2-character-roles narrative, then compute the share of each combinations with other character-roles. One interpretation is that the lower the share displayed on the diagonal, the more complex discussions involving that character-

[4] Note that we are not capturing the most recent climate bills in the U.S., which might reinstate more positive narratives about the government's role.

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



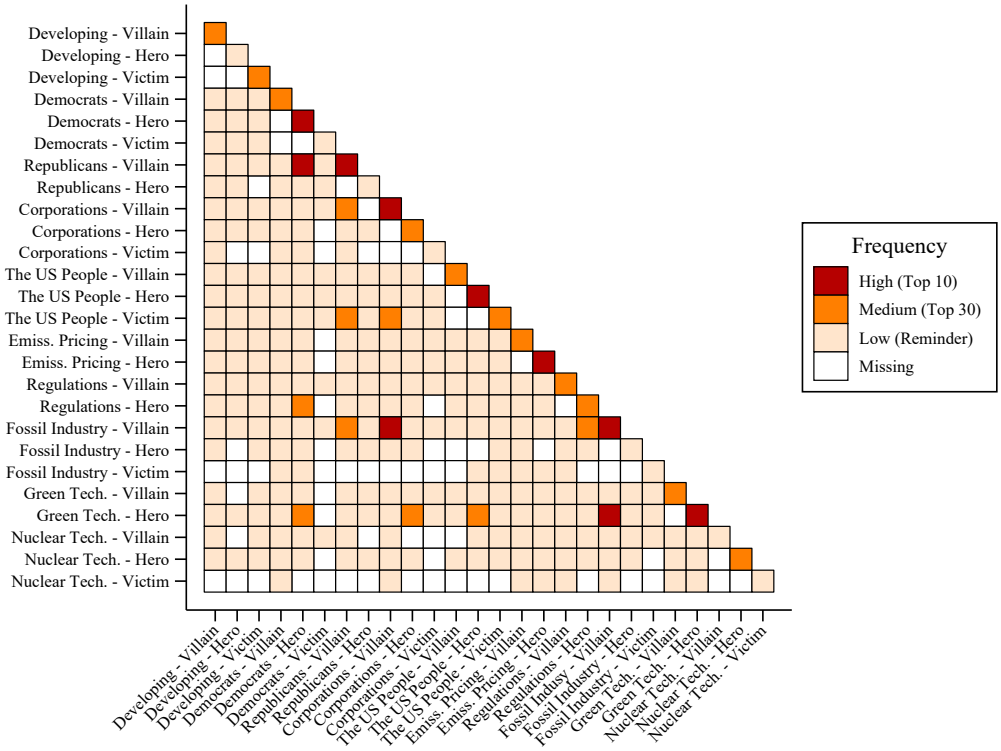
Notes: The figure shows the frequency of each character-role over time, in the dataset of relevant tweets used in our analysis. We define a tweet as relevant if it features at least one character from our list. We include only character-roles that appear at least 100 times, thus excluding U.S. Republicans-victim, Emission Pricing-victim, Regulations-victim, and Green Tech-victim. The shares of character-roles are cumulative, i.e. sum up to 100, and are computed on a yearly basis. We label the most common fifteen character-roles. In parentheses, each label indicates the share of the corresponding character-role in the first period, the average share across all periods, and the share in the last period, in this order. The non-labeled character-roles are the following, from top to bottom of the graph: Developing Economies-hero, U.S. Democrats-victim, U.S. Republicans-hero, Corporations-hero, Corporations-victim, U.S. People-villain, Fossil Industry-hero, Fossil Industry-victim, Nuclear Tech-villain, Nuclear Tech-hero, and Nuclear Tech-victim.

role. Shares are not displayed numerically to avoid visual overload, but we indicate the top 5 and top 30 most frequent, with a color scheme.

For the majority of character-roles, the simple narrative form is the most common way in which they appear. In the matrix, we highlight the most frequent co-occurrences. All top 5 most frequent narratives are single character-roles narratives being U.S. Democrats-hero, U.S. Republicans-villain, Corporations-villain, Fossil Industry-villain and Green Tech-hero. Nevertheless, several combinations of narratives are very common in our dataset among which U.S. People-victim that appears paired to Corporations-villain. Many frequent pairs are constructed by associating Green Tech-hero to other character-roles among which U.S. People-hero. There are also combinations that never occur together, like Fossil Industry-hero and Emission Pricing-hero.

Although Twitter is dominated by short and relatively simple narratives, likely due to the platform's word limit, it is important to note that complex narratives – combining several character-

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



Notes: The figure shows how often each character-role appears alone or alongside another character-role in relevant tweets that contain one or two character-roles. 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 U.S. Republicans-victim, Emission Pricing-victim, Regulations-victim, and Green Tech-victim. Frequencies are computed by dividing the number of times a particular pair or single character-role appears by the total number of tweets containing 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 do not display exact shares to avoid visual overloading. We use a color scheme to highlight the top 10 most frequent 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 are in order: Green Tech-hero (10.27%), U.S. Republicans-villain (8.95%), Fossil Industry-villain (5.56%), U.S. Democrats-hero (4.21%), Corporations-villain (3.66%), U.S. People-hero (3.40%), Fossil Industry-villain + Corporations-villain (3.27%), Green Tech-hero + Fossil Industry-villain (3.02%), Emission Pricing-hero (2.40%), U.S. Republicans-villain + U.S. Democrats-hero (1.92%).

roles – play also a significant role in the discourse on climate policy. Our approach is effective in capturing both simple, slogan-like narratives and more articulate narratives featuring multiple character-roles. Although our method may sacrifice some precision in capturing the exact causal links and direction of the narrative conveyed in the text, it enables the analysis of large amounts of data and allows for a nuanced depiction of the discussion.

5 Main Results: Virality of Political Narratives

Access to information has reached unprecedented levels, nevertheless the impact of facts is often overshadowed by the power of narratives (Bursztyn et al. 2023b). Narratives do not merely describe

facts; they shape perceptions, opinions, and behaviors, often leading to measurable economic and social consequences (Andre et al. 2025). One of the primary mechanisms through which narratives gain traction and influence is virality, particularly on social media (Shiller 2017). Understanding what makes narratives go viral is therefore crucial to assessing their broader societal impact. In this section, we analyze the virality of narratives, first examining the fundamental nature of virality on Twitter, then exploring the relationship between narratives and virality, and finally identifying the key determinants that drive a narrative's virality.

5.1 Virality of Character-Roles

We define virality by the number of retweets a tweet receives. The first research question we address is straightforward: how viral are the political narratives in our dataset? We exploit the flexibility and capacity of our framework, and analyze virality for all combinations of one or two character-roles in Figure 5. The squares of the matrix capture combinations of character-roles, and the color of the squares indicates the level of retweets obtained by the specific character-role combination, out of the total amount of retweets obtained by any narrative containing one or two character-roles. Political narratives with single character-roles are shown on the diagonal, complex narrative combinations in the other entries of the matrices. The matrix is symmetrical.

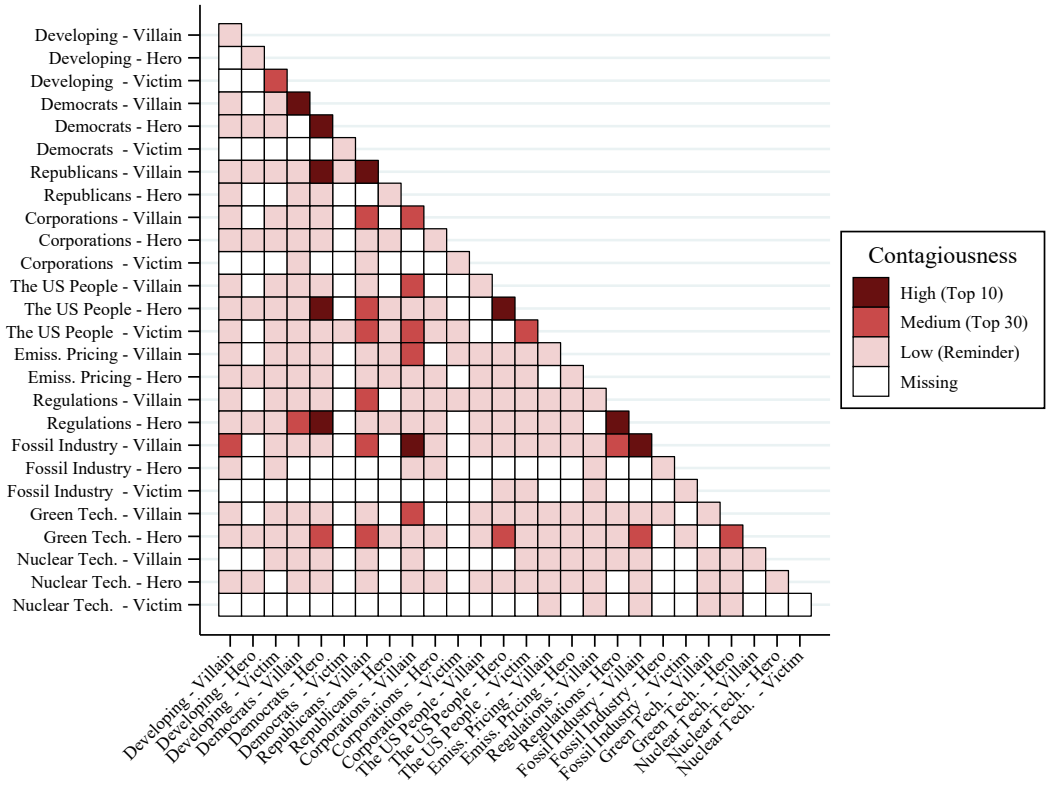
Certain character-role combinations prove especially viral. Among the five most retweeted narratives, the most common pattern sees U.S. Democrats as heroes and U.S. Republicans as villains. Narratives that portray U.S. People as heroes likewise attract high retweet counts, whether they appear alone or alongside other roles. Depictions of the Fossil Industry as a villain also spread widely, particularly when Corporations are portrayed alongside as villains, a pairing that itself ranks in the top five. Although narratives featuring human actors tend to dominate, non-human roles such as Regulations and Green Tech still break through, appearing in several of the thirty most retweeted narratives.

While Figure 5 provides a descriptive overview of the average virality of political narratives, it does not speak to the difference between the virality of political narratives, versus that of tweets only featuring characters in a neutral way. Figure 6 provides descriptive insights into this distribution. It displays the Log-Log Rank Distribution of retweets, distinguishing between tweets that feature at least one character-role from those that only feature characters in a neutral way.

In the figure, the x-axis represents the logarithm of the rank of tweets based on their retweets, respectively, with rank 1 corresponding to the most retweeted tweet in the dataset. The y-axis represents respectively the logarithm of the number of retweets, plus one. The distribution reveals key insights. First, for both neutral tweets and tweets containing a character-role, the distribution takes a linear shape. This suggests that virality follows a power-law rank distribution among the tweets in our dataset, which can be represented by the following functional form:

$$R(x) \propto x^{-\beta} \tag{1}$$

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

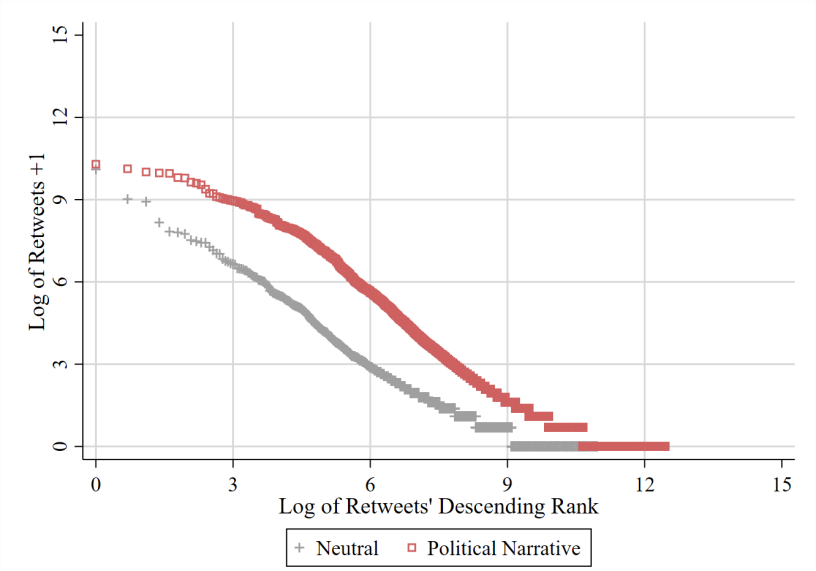


Notes: The figure shows the retweet rate of each character-role that appears alone or alongside another character-role in relevant tweets that contain one or two character-roles. We define a tweet as relevant if it features at least one character from our list. We include only character-roles that appear at least 100 times, thus excluding U.S. Republicans-victim, Emission Pricing-victim, Regulations-victim, and Green Tech-victim. Retweet rates are computed by dividing the total number of retweets received by a particular pair or single character-role by the total number of retweets of tweets containing one or two character-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. 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: U.S. Republicans-villain (16.72%), U.S. Democrats-hero (12.98%), U.S. People-hero (7.92%), Fossil Industry-villain + Corporations-villain (7.89%), U.S. Republicans-villain + U.S. Democrats-hero (6.51%), U.S. Democrats-villain (4.79%), Fossil Industry-villain (3.07%), U.S. People-hero + U.S. Democrats-hero (2.70%), Regulations-hero (2.57%), Regulations-hero + U.S. Democrats-hero (2.51%).

where $R(x)$ represents the expected number of retweets at rank x, and β is the scaling exponent that determines how quickly engagement declines as rank increases. A higher value of β results in a steeper decline, meaning engagement is concentrated in a few highly viral tweets, while most receive minimal interaction.

The exponent β directly influences the slope of the curves in the log-log distribution graphs. A steeper slope (β large) indicates a rapid drop from the most retweeted tweets to the least, suggesting that engagement is highly concentrated among a few viral tweets. Conversely, a flatter slope (β small) implies a slower decline, meaning engagement is more evenly distributed across tweets. The vertical position of the curves, at the same rank, provides additional information: at parity of rank,

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



Notes: The figure provides insights into the nature of virality in the tweets from our sample. It displays the log-log rank distribution of retweets for relevant tweets. 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 U.S. Republicans-victim, Emission Pricing-victim, Regulations-victim, and Green Tech-victim. The x-axis represents the logarithm of the rank distribution, where rank 1 corresponds to the most retweeted tweet in the sample. The y-axis represents the logarithm of the number of retweets, plus one. The slope of the curve reflects how rapidly the dataset transitions from the most viral tweets to the least viral. The vertical position of the curve at a given rank provides a measure of relative virality—tweets with a higher curve position receive more engagement than those at the same rank in a dataset with a lower curve.

a higher-positioned curve signals a greater level of engagement, meaning that tweets in this subset receive more retweets or likes compared to those in another subset with a lower-positioned curve. This distinction is particularly relevant when comparing narrative tweets with neutral ones. The figures show how the distribution for tweets with narratives is vertically shifted, indicating that at parity of rank narrative tweets are generally more viral.

5.2 The Determinants of Virality

Beyond more descriptive results, we also use a regression framework to analyze the determinants of virality more generally. We use the following Poisson pseudo-maximum likelihood regression equation:

$$E[Y_{i,s,t}|D_{i,s,t}] = exp[\alpha + \beta D_{i,s,t} + \theta X_{i,s,t} + \gamma_t + \delta_s] \quad \forall i \in C, \tag{2}$$

where $Y_{i,s,t}$ refers to the count of retweets for tweet i, originating from state s in year t. Each tweet i in the model is such that $i \in C$, where C is the sample of tweets located in the U.S. with precision at least at the state level. α is a constant term, while γ_t and δ_s refer respectively to the

year and state fixed effects. $D_{i,s,t}$ refers to a dummy variable reflecting the comparison of interest. E.g., when comparing villain and hero narratives it equals 1 if a villain narrative is present in the tweet and 0 for hero narratives. We exclude tweets that feature neither of the two, as well as all tweets that feature both simultaneously. $X_{i,s,t}$ collects the number of hashtags, mentions and words in the tweet as well as the number of followers, following and tweets ever produced by the user. We cluster standard errors at the week level.

Before turning to estimates, it is useful to clarify how we read $\widehat{\beta}$ in (2) in light of potential omitted drivers of virality on the platform. Let the true process be

$$Y_{i,s,t} = \exp\{\alpha + \beta D_{i,s,t} + \gamma A_{i,s,t} + \theta' X_{i,s,t} + \gamma_t + \delta_s + u_{i,s,t}\},$$

where $A_{i,s,t}$ denotes omitted variables, e.g. unobserved ranking/moderation shifters employed by Twitter. Estimating (2) without $A_{i,s,t}$ yields the standard omitted-variable term

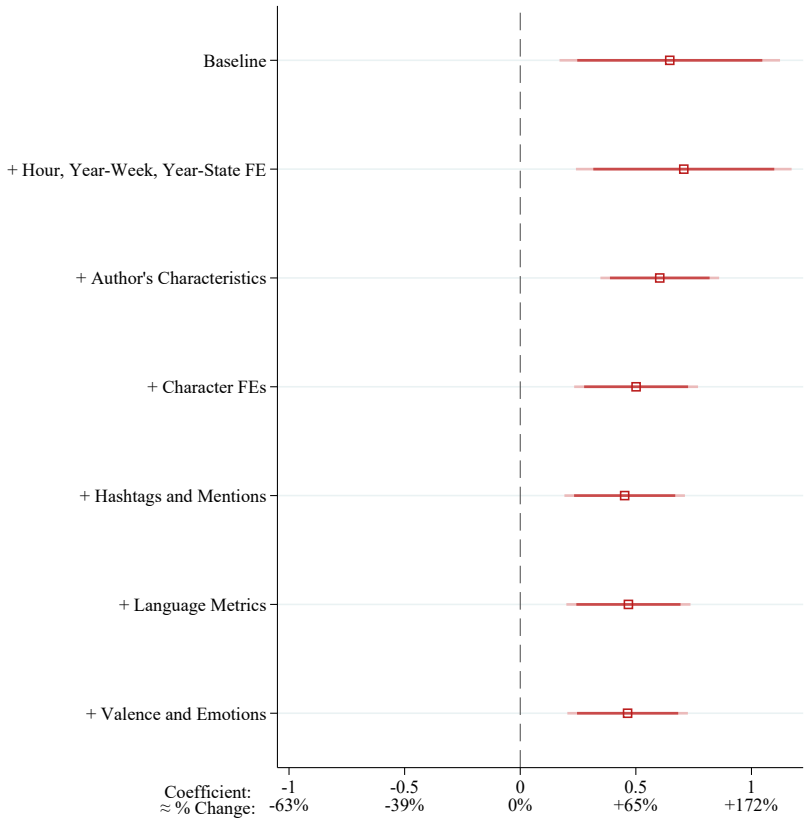
$$\gamma \cdot \frac{(D_{i,s,t}, A_{i,s,t} \mid X_{i,s,t}, \gamma_t, \delta_s)}{(D_{i,s,t} \mid X_{i,s,t}, \gamma_t, \delta_s)}$$

in $\widehat{\beta}$. If narrative tweets are systematically surfaced more ($\gamma > 0$ and $(D, A \mid \cdot) > 0$), $\widehat{\beta}$ is biased upward; if moderation disproportionately suppresses some role uses so that $(D, A \mid \cdot) < 0$, the bias is downward. This logic does not require time-variation in $A_{i,s,t}$; a time-invariant ranking rule can still generate bias whenever it differentially favors $D_{i,s,t} = 1$ tweets. In addition, misclassification in our role labels $D_{i,s,t}$ that is approximately mean-zero and independent of $Y_{i,s,t}$ conditional on covariates acts like classical measurement error, attenuating estimates toward zero under a linear approximation and, analogously, weakening PPML coefficients. Taken together, engagement-based surfacing can push $\widehat{\beta}$ upward, while moderation and measurement error plausibly pull it downward; the net bias is a priori ambiguous.

Our empirical design mitigates the most likely channels for $(D_{i,s,t}, A_{i,s,t} \mid \cdot)$ by conditioning on rich author and time effects and on text features—sentiment, emotions, length, hashtags, and mentions—that ranking systems can observe or proxy. Because the use of role language and emotional tone often co-occur within a tweet, conditioning on these features is appropriate rather than over-controlling. With these safeguards in place, we interpret $\widehat{\beta}$ as the association between character–role assignment and virality over and above observable author and text characteristics, and we place emphasis on within-sample contrasts across roles (hero, villain, victim versus neutral) and extensive robustness. We now turn to the estimates.

We begin our analysis by addressing the most basic question: Does using political narratives impact virality at all? Figure 7 provides insight into this question through a coefficient plot illustrating the impact of political narratives on virality compared to tweets where also at least one character appear, but only in neutral roles. The plot is structured top to bottom estimating a sequence of increasingly stringent models, beginning with a baseline specification without controls. The second model introduces hour and year-state fixed effects (FEs). The third model accounts for author characteristics, including whether the account is verified, number of followers and follow-

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



Notes: The figure shows the coefficients of a Poisson Pseudo-Maximum Likelihood 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. 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 U.S. Republicans-victim, Emission Pricing-victim, Regulations-victim, and Green Tech-victim. 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 baseline 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 model controls for tweet characteristics (number of hashtags and mentions). The sixth and seventh models include, respectively, language metrics and valence/emotions, as used in the descriptive analysis above.

ings, total tweets created, and whether the author identifies as Democratic or Republican, religious, highly educated, or as a parent. The fourth model further controls for character fixed effects, and the fifth model incorporates tweet characteristics, such as word count, number of hashtags, and number of handles. The final column controls for emotions and valence.

Narratives have a persistent and positive impact on virality, an effect that remains significant even as we introduce increasingly stringent model specifications. One key factor influencing virality is the author's characteristics, which reduce the estimated effect size while simultaneously improving model precision by reducing noise. A natural concern is whether the observed effect is simply capturing the virality of the characters included in our analysis. However, even after introducing

character fixed effects, the impact of narratives on virality remains substantial, suggesting that the effect is not merely driven by the presence of specific characters.

As discussed earlier, multiple stylistic text elements are able to construct political narratives, including tone, plot structure, character roles, and emotional framing. In previous sections, we examined how narratives differ from neutral tweets in terms of valence, emotional content, and text quality. While narratives exhibit clear distinctions along these dimensions, controlling for these factors does not eliminate their effect on virality. This suggests that narratives influence engagement through mechanisms beyond just emotion or linguistic style, reinforcing their distinct role in shaping public discourse.

Our definition and detailed, structured measurement of political narratives then allows us to refine our analysis by moving beyond the overall impact of featuring any narrative and addressing a more nuanced question: Which roles of the drama triangle drive virality most effectively, and do all roles contribute equally? We answer this question in Table 3, where present the effect of featuring specific character roles on virality.

To ensure a reliable and mutually exclusive comparison, the results focus solely on narratives containing a single character-role. We compare the effect of complex narrative in the next part of the analysis. Column 1 of the table reports the effect of featuring a hero narrative compared to neutral tweets, column 2 compares villain narratives to neutral tweets, and column 3 does the same for victim narratives. Since all samples are mutually exclusive, the estimated effects reflect the distinct impact of each role. Columns 4 through 7 introduce models that include all roles except for one, allowing for direct comparisons between them. In these models, the reference category changes across columns: column 4 uses neutral tweets as the baseline, column 5 hero narratives, 6 villain, and column 7 uses as baseline victim narratives.

Several noteworthy patterns emerge. Hero and villain narratives significantly enhance virality compared to tweets where characters appear in neutral roles. Column 1 shows that hero narratives increase retweets by approximately 55%, while villain narratives lead to a much stronger increase of 170% (computed as $\approx e^{\beta} - 1$ for Poisson models). In contrast, victim narratives do not exhibit a statistically significant effect on virality compared to neutral-character tweets, suggesting that they do not drive engagement in the same way. When considering the full models in columns 4 through 7, a more nuanced pattern emerges. Villain narratives consistently stand out, surpassing both hero and victim narratives in driving virality, reinforcing the idea that narratives centered around opposition and conflict tend to generate the highest levels of engagement. These findings suggest that negative framing, conflict-driven narratives, and oppositional storytelling play a crucial role in determining virality on social media.

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

Dependent Variable	**Retweets' Count**						
	Coeff./SE/p-value						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hero	0.445			0.412		-0.552	0.239
	(0.193)			(0.211)		(0.210)	(0.279)
	[0.021]			[0.051]		[0.009]	[0.393]
Villain		0.923		0.964	0.552		0.791
		(0.135)		(0.198)	(0.210)		(0.291)
		[0.000]		[0.000]	[0.009]		[0.007]
Victim			0.090	0.173	-0.239	-0.791	
			(0.326)	(0.287)	(0.279)	(0.291)	
			[0.783]	[0.546]	[0.393]	[0.007]	
Neutral					-0.412	-0.964	-0.173
					(0.211)	(0.198)	(0.287)
					[0.051]	[0.000]	[0.546]
Sample: Neutral	✓	✓	✓	✓	✓	✓	✓
Sample: Hero	✓			✓	✓	✓	✓
Sample: Villain		✓		✓	✓	✓	✓
Sample: Victim			✓	✓	✓	✓	✓
Mean Outcome Reference Group	1.972	1.972	1.972	1.972	4.350	3.683	2.395
Pseudo R2	0.79	0.70	0.68	0.71	0.71	0.71	0.71
Obs	77047	90057	43952	137664	137664	137664	137664

Notes: The tables show the coefficients of Poisson Pseudo-Maximum Likelihood regressions testing the effect of featuring Hero, Villain, and Victim roles on virality, measured as the count of retweets, Panel A, and likes, Panel B. The regression models include relevant tweets featuring at most one character, either neutral or in a role. 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 U.S. Republicans-victim, Emission Pricing-victim, Regulations-victim, and Green Tech-victim. The regressions' coefficients can be transformed to percentage change with the formula: $\approx e^{\beta} - 1$. All regressions include authors' characteristics (verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status), and character fixed effects. We include hour, week of the year and year-state fixed effects. Standard errors are clustered at the week level, covering all weeks in our time frame.

5.3 Heterogeneous Effects

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 begin by addressing the following question: Does narrative complexity matter for virality? We define a narrative as simple if it contains only one of the three roles in the Drama Triangle – even if multiple characters are present within that role. Conversely, we define a narrative as complex if it features two or all three roles within the same tweet. This classification reflects the underlying structure of the message. When multiple characters appear but share the same role, the tone, emotional framing, and intended opinion remain consistent. However, when multiple roles are present, the narrative becomes more nuanced, incorporating multiple perspectives or conflicting viewpoints, thus increasing its complexity.

Figure 8 provides insights into the effect of narrative complexity on virality. We estimate

models where each single-role narrative (e.g., hero) is compared to its three complex alternatives (e.g., hero-villain, hero-victim, and hero-villain-victim), as well as to a reinforced simple narrative, where multiple characters appear within the same role (e.g., multiple heroes). Panel A of the figure presents results for hero narratives, Panel B for villain narratives, and Panel C for victim narratives.

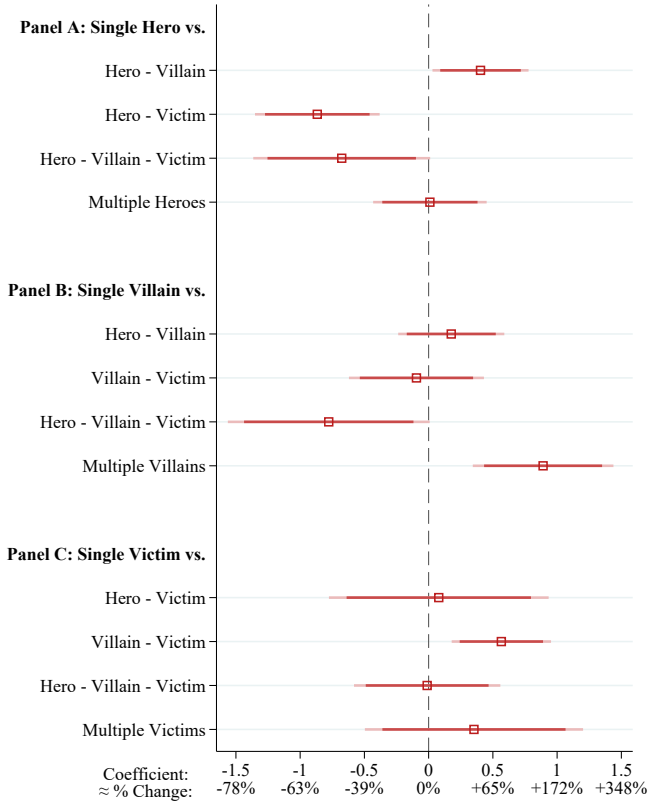
Compared to tweets featuring a single-character hero narrative, increasing complexity by introducing a hero-victim or hero-victim-villain structure reduces overall virality. However, when the hero role is combined exclusively with a villain, contagiousness increases, while popularity remains unchanged. A similar pattern emerges for villain narratives, where complexity tends to reduce virality, particularly when the full Drama Triangle is present. Notably, when villains are reinforced within a simple narrative – meaning multiple characters share the villain role – virality increases even further, suggesting that narratives centered around multiple villain tend to gain more traction. For victim narratives, no consistent pattern emerges, except when a victim is paired with a villain, which significantly boosts contagiousness.

Overall, these findings suggest a clear and consistent pattern: narrative complexity tends to reduce virality. When multiple character-roles interact within the same tweet, the added nuance appears to be detrimental to both retweets and likes. In contrast, when a message is reinforced by multiple characters of the same role, virality increases. This effect is particularly pronounced for villain narratives, which emerge as a strong driver of engagement. The contrast between villains and either heroes or victims appears to amplify a narrative's impact, making it more compelling. Most strikingly, tweets featuring multiple villains generate the highest levels of virality, suggesting that oppositional framing and conflict-driven narratives are particularly effective in capturing attention. Do people engage most with what they oppose?

In the final part of this section we ask whether certain kinds of narratives spread more easily than others. We exploit two dimensions of our narrative framework and classification pipeline. First, we test whether the type of characters matters for virality, comparing narratives built around human actors (e.g., U.S. People or Corporations) with those centered on instrumental actors such as regulations or technologies. Second, we test whether the number of character-roles in a tweet makes a difference, does adding more characters framed as heroes, villains, or victims make a difference for virality? Table 9a and Figure 9b show the results of our heterogeneous-effects analysis.

Table 9a reports Poisson Pseudo-Maximum Likelihood estimates, where the dependent variable is the number of retweets each tweet receives. Columns (1) and (2) compare tweets that feature only human characters with tweets that feature only instrumental characters. To make the comparison clean, we restrict the sample to tweets that include at least one character-role and place each tweet in one of two mutually exclusive groups: all-human or all-instrumental. Column (1) narrows the analysis to tweets with a single character-role, while Column (2) allows multiple roles. Columns (3) and (4) focus on the number of roles. Column (3) includes the raw count of character-roles as a regressor; Column (4) adds the squared term to test for non-linear effects. Figure 9b complements these regressions by plotting the marginal change in expected retweets associated with each additional character-role, providing a more intuitive picture of how narrative complexity influences

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



Notes: The figure shows the coefficients from Poisson Pseudo-Maximum Likelihood regressions testing the effect of character-role combinations on virality. 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 U.S. Republicans-victim, Emission Pricing-victim, Regulations-victim, and Green Tech-victim. The dependent variable is the tweet-level count of retweets. The x-axis reports coefficient estimates along with the corresponding approximate percentage change, computed as follows: $\approx e^{\beta} - 1$. Each panel corresponds to a single regression model. Panel A examines the effects for hero, with the comparison group being tweets featuring a single hero character. The regressors capture character-role combinations where the hero is paired with others. Specifically, we estimate the effects of a hero paired with a villain or more villains, a hero paired with a victim or more victims, a hero paired with both villains and victims, and tweets featuring multiple heroes. Panels B and C follow the same structure, focusing on villain and victim, respectively, with their corresponding comparison groups. All models include hour, week of the year, and year-state fixed effects. Standard errors are clustered at the week level, covering all weeks in the time period.

virality.

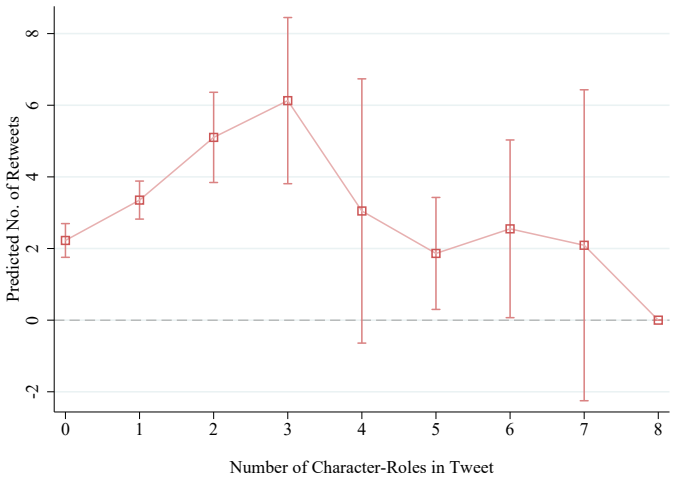
The results paint a consistent picture. Tweets built around human characters spread much farther than those built around instrumental characters such as regulations or technologies. This holds whether a tweet features a single role or several, and the effect sizes are large: shifting from an instrumental to a human focus raises expected retweets. In short, solutions-oriented or policy-focused messages attract far less engagement than stories that center on people. Turning to narrative complexity, adding characters boosts virality at first but gives diminishing returns. Moving from one to two character-roles delivers the biggest jump in retweets, and a third role still

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

(a) Heterogeneous Effects

Dependent Variable	**Retweets Count**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Human vs. Instrument	0.712	1.937		
(mut. excl.)	(0.341)	(0.861)		
	[0.037]	[0.024]		
Sum of CRs			0.214	0.659
			(0.057)	(0.150)
			[0.000]	[0.000]
Sum of CRs Squared				-0.125
				(0.045)
				[0.006]
Single Character-Role	✓			
Multiple Character-Roles		✓	✓	✓
Mean Outcome Reference Group	12.201	12.201	10.923	10.923
Pseudo R2	0.71	0.80	0.62	0.62
Obs	142611	42951	309531	309531

(b) Marginal Effect of Adding Character-Roles



Notes: The exhibit provides an overview of the heterogeneous effects of narratives on virality. The table reports Poisson Pseudo-Maximum Likelihood estimates where the dependent variable is the number of retweets a tweet receives. Columns (1)–(2) isolate character type: the sample is limited to tweets that contain at least one character-role and each observation is placed in a mutually exclusive group of either (i) tweets featuring only human characters or (ii) tweets featuring only instrumental characters. Column (1) restricts the comparison to tweets with a single character-role; Column (2) allows multiple roles within the tweet. Columns (3)–(4) examine narrative complexity: Column (3) includes the total count of character-roles as a regressor, while Column (4) adds the squared term to test for non-linear effects. Standard errors are robust to heteroskedasticity. Figure 9b complements the regression results by plotting the marginal change in expected retweets associated with each additional character-role, providing an intuitive visualization of how increasing narrative complexity affects virality.

helps, but the incremental gain is small by the time a fourth role appears. Beyond that point the curve flattens, suggesting that audiences lose interest, or find the story harder to process, once the cast grows too crowded.

6 Experimental Results: Beliefs, Preferences and Memory

Our analysis identifies what drives virality in social media. Using a large-scale observational dataset of tweets, we find that present characters within a narrative consistently go more viral than those that describe the same characters in neutral terms. But not all narratives are equally effective. Hero and villain narratives outperform victim narratives, and tweets that feature villains generate twice as much engagement as those that feature heroes. Including both heroes and victims alongside a villain further boosts virality. However, tweets with multiple heroes or villains perform better than those that combine all character types, suggesting that complexity may reduce engagement, while clear and reinforced character roles enhance it. The next step in our analysis is to examine the impact of narratives on beliefs and preferences.

The next step in our analysis examines whether narratives do more than capture attention. Do narratives also influence beliefs and preferences? While we identify which narratives go viral, it remains an open question whether mere exposure to a viral narrative changes how people think

or what they support. To test this, we run three parallel survey experiments, each focused on a different narrative.

We define the impact of narratives along two dimensions. First, narratives may shape beliefs by influencing how individuals form expectations about the future. Beliefs matter because they guide how people interpret information and make decisions, serving both psychological and functional roles (Bénabou, Falk, and Tirole 2020). Second, narratives may shift preferences by changing attitudes toward specific policies. We view belief change as a necessary condition for preference change: individuals are unlikely to revise their policy positions unless their expectations evolve. However, research in psychology and economics shows that changes in beliefs do not always lead to preference shifts, as preferences often reflect deeper elements of identity, personal experience, and social norms (Akerlof and Kranton 2000).

We further distinguish between stated and revealed preferences. Survey responses about policy support offer one measure of preferences, but they often fail to predict real-world behavior. Individuals may express attitudes that do not translate into action. To address this, we test whether narratives influence revealed preferences, measured through decisions involving real stakes. Behavioral choices tied to actual incentives provide a more reliable indicator of preferences than self-reported attitudes. Finally, building on the recent work by Graeber, Roth, and Zimmermann (2024), we also explore the impact narratives may have on memory retention, a potential mechanism how narratives may influence belief and preferences changes.

The remainder of the section is organized as follows. First we provide an overview on the design of our experiments in Section 6.1, second we show our main results in Section 6.2. Finally, we explore the potential impact of narratives on memory in Section 6.3.

6.1 Experimental Design

To assess the impact of narratives on beliefs and preferences, we conduct three pre-registered online experiments, each following an identical design. Our goal is to test the effect of being exposed to a narrative tweet compared to a neutral tweet that presents the same factual information and characters but without narrative framing. Building on our earlier findings about narrative virality on Twitter, we focus on specific character-role combinations that were particularly viral during our observation period. We select three such narrative structures and test each in a separate experiment.

In the first experiment, the narrative tweet frames Green Tech and the U.S. 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. In all three experiments, renewable technologies are framed as heroes. We anchor each narrative around this pivotal character-role to ensure comparability and to facilitate pooled analysis across experiments. 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. Balance tests confirm that the treatment and control groups are comparable, as we show in the Appendix Figure F.1. Each experiment includes a follow-up survey administered one day later. In this second survey, participants are shown the same feed as before, but with the third post blurred. They are asked to recall both the factual content and the characters mentioned in that post.

To measure the impact of narratives, we design questions that capture changes in beliefs, preferences, and memory. First, we assess whether narratives influence beliefs by asking participants to predict the share of renewable energy in the U.S. by 2035. Next, we examine policy preferences by measuring support for specific policies related to the characters in the tweets. We evaluate revealed preferences through a dictator game in which participants decide how to allocate $25 between themselves and a green technology organization. This decision is incentivized through a lottery, allowing us to observe actual financial trade-offs rather than hypothetical choices. Finally, we measure memory retention by encoding an open question that we pose on the day itself and the day after, asking the participants to mention anything they remember about the tweet.

6.2 Experimental Results: Beliefs and Preferences

We begin by verifying that participants in the treatment group perceived the tweets as narratives – that is, that they recognized the hero–villain–victim framing. This check is possible because, while neutral posts mention characters without assigning them a role, narrative posts explicitly build the story by portraying these characters as heroes, villains, or victims. Hence, as pre-registered, we test whether the narrative manipulation was effective by encoding the answers of the following open question asked 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."*.

Table 4 presents the results of our manipulation check. We assess whether participants in the narrative treatment were more likely to mention characters in a specific role and mention causal links between them. We show results separately for each experiment, Hero-hero, Hero-hero-villain, and Villain-villain-hero, as well as for all experiments combined, where we include experiment fixed effects (FEs). In columns (1) to (4) we show linear OLS models where the dependent variable is 1 if the participant mentions the character-role, and 0 otherwise. In columns (5) to (8) we show the same but conditionally on the participants mentioning the character at all in their answer. The results are highly consistent across specifications: Narratives affect their perception of character-roles, with treatment being roughly 30% more likely to mention the hero-villain-victim framing in

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

Dependent Variable	**Mention of Character-Role**							
	Coeff./SE/p-value							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Political Narrative vs. No Role	0.293 (0.032) [0.000]	0.281 (0.032) [0.000]	0.254 (0.031) [0.000]	0.271 (0.018) [0.000]	0.358 (0.034) [0.000]	0.322 (0.035) [0.000]	0.280 (0.032) [0.000]	0.316 (0.019) [0.000]
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Outcome | 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	2931	742	686	695	2123

Notes: The table displays the coefficients of OLS regression models providing a manipulation check of the effectiveness of the narrative treatment. The check is done by exploring whether participants are more likely to mention the character-roles, defining feature of a narrative, when exposed to the narrative treatment rather than to the control. Columns 1, 2, 3, and 4 display the effect of being exposed to the narrative on the likelihood of mentioning the character-role (hero, villain, victim). Columns 5, 6, 7, and 8 display the same conditional on mentioning the character at all. The outcome variable is always a dichotomous indicator of whether the participant mentions the character roles when asked to recall anything about the text they saw. All models include income, education, political preference, age, and sex as controls. We use robust standard errors.

their answers in all specifications.

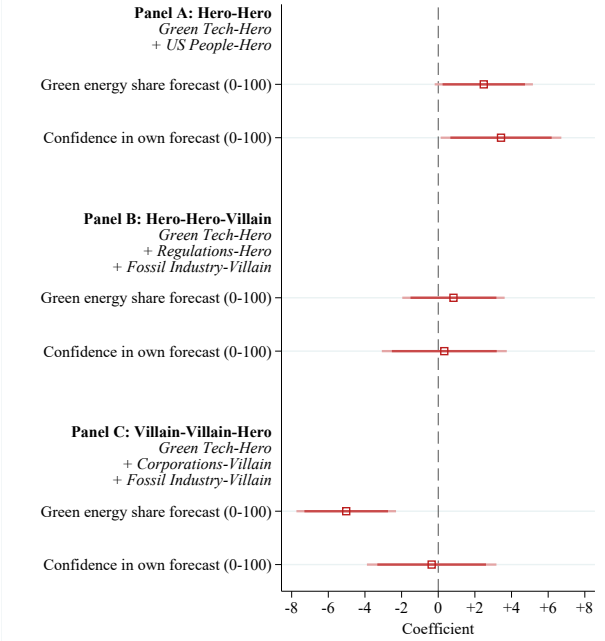
Once we established that the narrative treatment effectively shaped participants' perceptions, we turn to its impact on beliefs and stated preferences. Figure 10 summarizes the results: the left panel (Figure 10a) shows effects on belief formation, while the right panel (Figure 10b) reports effects on stated policy preferences. As a reminder, we measure beliefs by asking participants what percentage of U.S. energy will come from green technologies in 2035, along with their confidence in that estimate. Stated preferences are captured through support for specific policies aligned with the heroic characters featured in the narratives.

Figure 10a breaks down the belief results by experiment – Hero–Hero (top panel), Hero–Hero–Villain (middle), and Villain–Villain–Hero (bottom). For each, we report the effect of the narrative treatment on both the belief (expected percentage) and the confidence in that belief. Several patterns emerge. The most positive narrative -— in the Hero–Hero experiment, where Green Tech and U.S. People are framed as heroes – leads to a small increase in predicted renewable energy share, though only the effect on confidence reaches statistical significance at the 5% level.

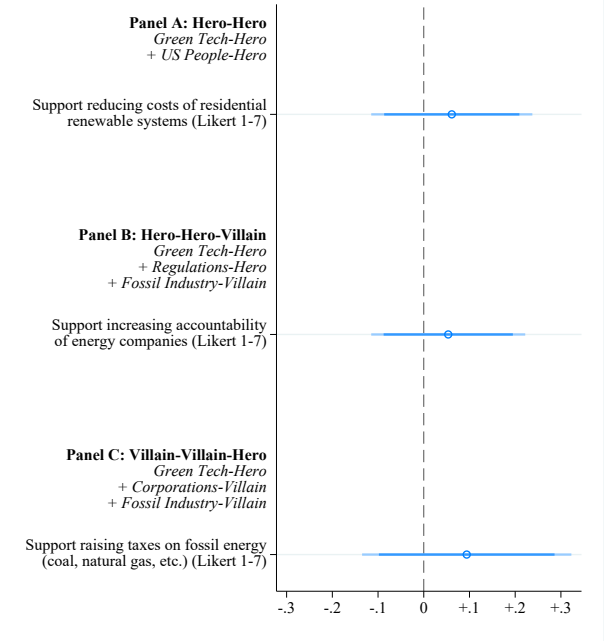
The second narrative, which pairs Green Tech and Regulations as heroes against the Fossil Industry as the villain, shows no clear effect on either beliefs or confidence. This recalls our earlier findings on virality: more complex or nuanced narratives may be less effective both in spreading online and in influencing attitudes. The strongest effect appears in the third experiment. In the Villain–Villain–Hero narrative, where the Fossil Industry and Corporations play as villains opposing Green Tech, participants predict a significantly lower share of renewable energy by 2035, about four percentage points less on average.

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

(a) Impact on Beliefs: Forecast for the Future



(b) Impact on Policy Preferences: Policy Support



Notes: The figures show the coefficients from OLS regression models analyzing the impact of the narrative treatment on beliefs and stated preferences. Panel A shows results for experiment Hero-Hero, Panel B for Hero-Hero-Villain and Panel C for Villain-Villain-Hero. In Figure 10a, for each experiment we show the impact of the narrative treatment on two outcomes: the participants' forecast, as an answer to the question *'What percentage of U.S. energy do you predict will come from renewable sources and green technology by the year 2035? Indicate a number between 0 and 100.'*, their confidence in the forecast, as answer to *'Your response on the previous screen suggests that by 2035, [x]% of U.S. 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]%?'*. Appendix Tables G.11 and G.2 show respectively the correspondent table of regressions, with and without controls. In Figure 10b, for each experiment 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. Appendix Tables G.12 and G.3 show respectively the correspondent table of regressions, with and without controls.

This finding aligns with a large literature in marketing, which shows that emotionally charged content is more likely to be shared and remembered (Berger 2011; Berger and Milkman 2012). Our results suggest that strong villain framing can heighten emotional engagement and lead to more pessimistic expectations about the future. In contrast, a narrative featuring only heroes does not shift expectations but does increase participants' confidence in their predictions. We interpret this as a reinforcement of their mental model, consistent with Graeber, Roth, and Zimmermann (2024): while beliefs may not change, narratives can still shape how strongly individuals hold them.

As discussed above, we view changes in beliefs as a precondition for a change in preferences. A single exposure to a narrative content has limited influence on belief formation, except in two cases: narratives that frame villains tend to lower expectations about the future, while those that highlight heroes increase participants' confidence in their existing beliefs. Given these modest shifts, we do

not expect substantial changes in policy preferences, and this is what we observe.

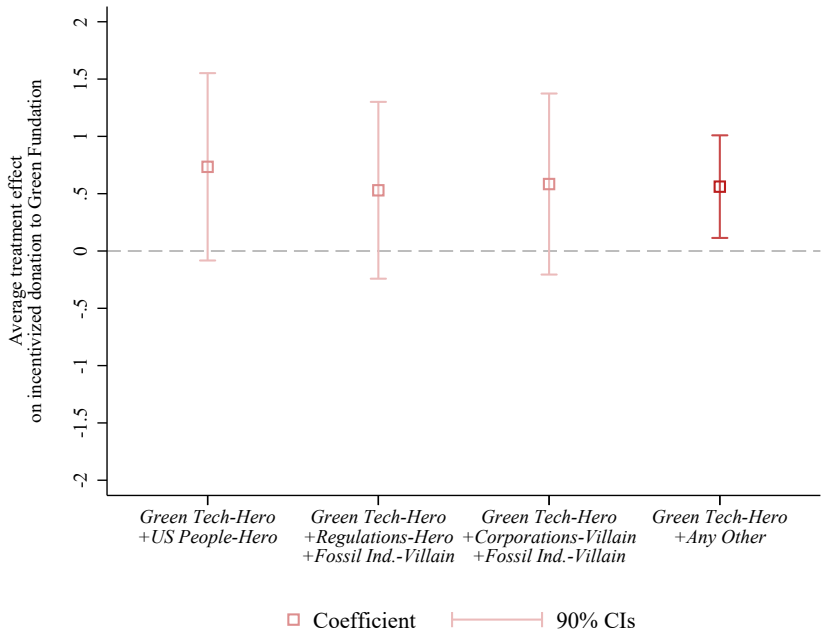
Figure 10b presents the results for stated preferences. The top panel shows results from the Hero–Hero experiment, where participants were asked to express support for reducing the cost of residential renewable energy systems. The middle panel reports results from the Hero–Hero–Villain experiment, which measured support for increasing accountability of energy companies. The bottom panel corresponds to the Villain–Villain–Hero experiment, where we asked participants about raising taxes on fossil energy sources.

Across all three experiments, exposure to the narrative treatments does not significantly alter stated preferences. This finding is consistent with existing research suggesting that policy preferences are more resistant to change than beliefs. However, in contrast to these null results, we observe meaningful shifts in revealed preferences, as shown in Figure 11. The figure displays coefficients from each individual experiment, along with a pooled estimate for our incentivized outcome: a dictator game in which participants allocate $25 between themselves and an organization that promotes green technology. This choice was incentivized through a lottery, enabling us to observe real financial trade-offs.

While statistical power is limited in these experiments, the results consistently point in the same direction. In all three treatments, and in the pooled analysis, exposure to the narrative increases participants' willingness to donate to the green technology institution. The pooled effect is statistically significant, and the magnitude of the coefficient is remarkably similar across the three experiments. This consistency is made possible by the shared presence of Green Technology as a heroic figure in all narrative treatments, which serves as a common anchor across conditions.

Our findings suggest that while narratives can subtly shift beliefs, their influence on expectations and policy preferences tends to be limited. Narratives reinforcing multiple villain roles can negatively affect future outlooks, while hero-focused narratives may boost confidence without altering expectations. These observations suggest that one-time exposure is insufficient to effect substantial change. Extensive research in psychology and marketing highlights the 'Mere Exposure Effect', where repeated exposure to a stimulus enhances familiarity and preference (Pohl 2022). This effect implies that continuous and repeated exposure to specific content, such as narratives, is necessary to influence beliefs and behaviors effectively. This points toward the crucial role of the virality of narratives; their impact is likely amplified when disseminated collectively and persistently across social media platforms.

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



Notes: The figure displays the coefficients and confidence intervals (90%) from OLS regression models analyzing the impact of the narratives on participants' revealed preferences. We measure revealed preferences with the decision to donate to a foundation 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. Moving from the left of the graph, coefficients 1, 2, and 3 show results for the single experiments, while the last coefficient on the right shows results for the pooled sample that uses all experiments and includes experiment fixed effects. All models include the following controls: income, education and politics. SEs are clustered at the participant level.

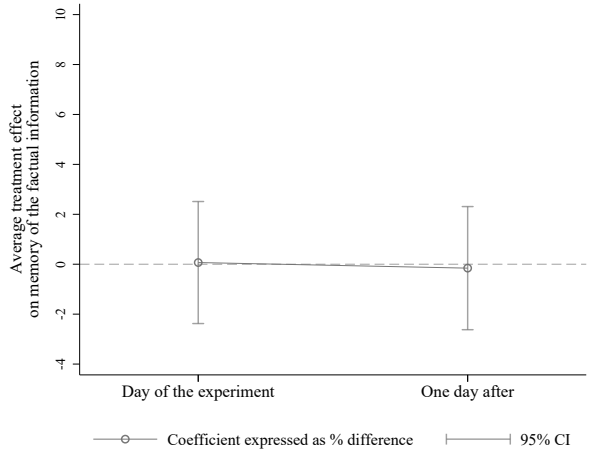
6.3 Experimental Results: Memory

While our previous results indicate that a single exposure to a narrative does not consistently shift beliefs or preferences, narratives may still play a crucial role in shaping how information is processed and retained. Specifically, narratives might act as a cognitive structuring tool, enhancing how individuals store and recall information. This could create the preconditions for future belief shifts, even if no immediate effect is observed.

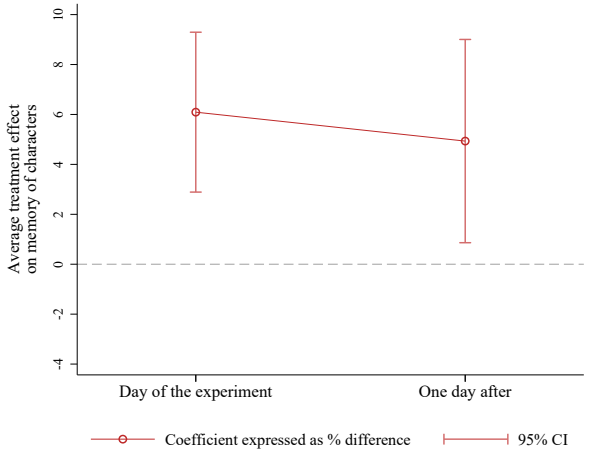
To investigate this, we included a memory recall task in our experiments. Participants were asked to recall two types of information: (1) the factual details presented in the narrative treatment and (2) anything else they remembered about the narrative treatment. This question was posed both on the day of the experiment and again in the follow-up survey conducted a day later. While narratives may not directly influence policy preferences, they reshape cognitive processing, making certain information more salient and retrievable over time. If narratives increase retention of key characters and causal links, they may function as a foundation upon which repeated exposure could later amplify changes in beliefs and shifts in attitudes.

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

(a) Memory of Factual Information



(b) Memory of Characters



Notes: The figures display the results from OLS regression models analyzing the impact of the narrative treatment on participants' memory. Figure 12a shows the effect on information retention, in the day of the main experiment, left coefficient, and a day later, right coefficient. Participants were asked to remember the factual information reported in the tweet, and the coefficients are expressed as percentage differences between treatment and control group. Figure 12b shows the effect on recalling the characters present in the text, the day of the experiment, on the left, and the day after the experiment, on the right. The dependent variable is obtained encoding an open-ended question that asked participants to recall anything from the text they saw, and coefficients are expressed as percentage differences between treatment and control group. All regressions pool all experiments together, include experimental wave FE, and the following controls: income, education and politics. SEs are clustered at the participant level.

Figure 12 provides insights into the effect of narratives on memory. The left panel shows the effect of being exposed to the narrative treatment on remembering the factual information contained in the tweet. The right panel shows the effect on remembering which characters were framed in the tweet. These results point to a very clear message. Narratives positively affect memory when it comes to remembering the characters and topic of it, while have a clear null effect when it comes to remembering the factual information featured.

7 Conclusion

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

By analyzing U.S. climate change policy discussions on Twitter over a decade, we show what makes narratives go viral. Political narratives, on average, are about sixty per cent more likely to be retweeted. This result holds controlling for a range of time and region fixed effects, for key authors characteristics like the number of followers, and when using character fixed effects. Negativity or emotions alone cannot explain that virality premium: when we hold eight discrete emotions and

continuous valence constant, the effect on virality is only slightly decreased and remains clearly significant. A villain framing lifts retweets by roughly 170 per cent, hero framing by about 55 per cent, while victim framing has little effect. Pairing another role with a villain raises virality, whereas adding other roles has an ambiguous effect, indicating that complexity can impose an attention cost. Human characters generally tend to go more viral than instrumental characters like technology or policies.

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

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

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

References

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

Akerlof, George A. and Rachel E. Kranton (Aug. 2000). "Economics and Identity". *Quarterly Journal of Economics* 115.3, pp. 715–753.

Akerlof, George A and Robert J Shiller (2010). *Animal Spirits: How Human Psychology Drives the Economy, and why it Matters for Global Capitalism*. Princeton University Press.

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

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

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

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

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

Ash, Elliott, Germain Gauthier, and Philine Widmer (2024). "RELATIO : Text Semantics Capture Political and Economic Narratives". *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". *The Economic Journal* 134.660, pp. 1331–1355.

Ash, Elliott et al. (2024a). "From viewers to voters: Tracing Fox News' impact on American democracy". *Journal of Public Economics* 240, p. 105256.

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

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

Barron, Kai and Tilman Fries (2023). "Narrative Persuasion". *CESifo Working Paper No. 10206*.

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

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

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

Berger, Jonah and Katherine L Milkman (2012). "What Makes Online Content Viral?" *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". *Social Psychology Quarterly* 81.3, pp. 228–247.

Bernauer, Thomas (2013). "Climate Change Politics". *Annual review of political science* 16, pp. 421–448.

Brittain, Melisa (2006). "Benevolent Invaders, Heroic Victims and Depraved Villains: White Femininity in Media Coverage of the Invasion of Iraq". *(En) Gendering the War on Terror*. Publisher: Hampshire, England and Burlington, VT: Ashgate, pp. 73–96.

Bursztyn, Leonardo et al. (2023a). "Justifying Dissent". *The Quarterly Journal of Economics* 138.3, pp. 1403–1451.

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

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

Caesmann, Marcel et al. (2024). "Censorship in Democracy". *arXiv preprint arXiv:2406.03393*. Publisher: University of Zurich.

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

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

Campante, Filipe, Ruben Durante, and Francesco Sobbrio (2018). "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation". *Journal of the European Economic Association* 16.4, pp. 1094–1136.

Chen, Jiafeng and Jonathan Roth (2024). "Logs with Zeros? Some Problems and Solutions". *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?" *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". *American Economic Review* 115.4, pp. 1258–1300.

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

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

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

Flynn, Joel and Karthik Sastry (2024). "The Macroeconomics of Narratives". *Working Paper*.

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

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

Gennaro, Gloria and Elliott Ash (Apr. 2022). "Emotion and Reason in Political Language". *The Economic Journal* 132.643, pp. 1037–1059.

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

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

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

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

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

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

Huber, Robert A, Esther Greussing, and Jakob-Moritz Eberl (2022). "From Populism to Climate Scepticism: The Role of Institutional Trust and Attitudes Towards Science". *Environmental Politics* 31.7, pp. 1115–1138.

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

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

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

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

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

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

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

Macaulay, Alistair and Wenting Song (2022). *Narrative-Driven Fluctuations in Sentiment: Evidence Linking Traditional and Social Media*. Economics Series Working Papers 973. MPRA Paper No. 113620.

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

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

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

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

Müller, Karsten and Carlo Schwarz (2023). "From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment". *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?" *Journal of Public Policy* 37.2, pp. 173–204.

O'Neill, Lachlan et al. (2021). *Quantitative Discourse Analysis at Scale-AI, NLP and the Transformer Revolution*. Tech. rep. SoDa Laboratories Working Paper Series No. 2021-12.

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

Pohl, Rüdiger, ed. (2022). *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment, and Memory.* Third edition. London New York: Routledge, Taylor & Francis Group.

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

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

– (2020a). *Narrative Economics: How stories go viral and drive major economic events*. Princeton University Press.

Shiller, Robert J (2020b). "Popular Economic Narratives Advancing the Longest US Expansion 2009–2019". *Journal of Policy Modeling* 42.4, pp. 791–798.

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

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

Verweij, Marco et al. (2006). "Clumsy Solutions for a Complex World: The Case of Climate Change". *Public Administration* 84.4. Publisher: Wiley Online Library, pp. 817–843.

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

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020). "Political Effects of the Internet and Social Media". *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 complementing 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 U.S. because Twitter plays a significant role in policy discussions, and 2010 marks the point when Twitter became a mainstream platform. While the sample of U.S. 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 more lose 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', 'fridaysforfutrue', '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 along 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-sensible 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

Three main datasets were used in this process. These datasets are stored in the folder '*output - data - processed_tweets - aggregated_data*' under the names *analysis_tweets_geolocalized_usa*, *tweets_visitrump_analysis*, and *tweets_visibiden_analysis*. The first dataset includes tweets collected from random days - one per month - and every Saturday from January 2010 to December 2021. These tweets were selected using specific keywords, as detailed in the corresponding Appendix of this paper. The second and third datasets contain tweets from users active in spreading climate policy narratives, particularly during the Trump and Biden elections. They include all relevant tweets from these users within the year surrounding each election, filtered by the keywords of interest used also to collect tweets for the previous dataset described above. The three datasets were concatenated, resulting in a final dataset containing 1.15 million tweets.

Annotation modality

The final concatenated dataset represent the final data used in this annotation process. For each tweet we query the OpenAI API, prompting the selected mode (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 want to know the following:

- 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.*'
- Relevant: 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.*'

In the second stage, if and only if a tweet is found to be 'relevant' at stage 1 - thus it is about climate policy and action - then the tweet goes through stage 2. Stage 2 consists in the individuation of characters and the classification of their role in the tweet. In particular, we query the model to find the following characters:

- 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
- US Democrat: 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 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
- 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

- 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
- 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)
- 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
- 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
- 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

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). Tweets not deemed "relevant" in stage 1 were classified simply as either 'irrelevant', 'assert', or 'deny' (see above for explanation).

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

Stage 2 prompt:

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

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

(0-4 as defined in step 2). - 'c': (0-4 as defined in step 2). - 'd': (0-4 as defined in step 2). - 'e': (0-4 as defined in step 2). - 'f': (0-4 as defined in step 2). - 'g': (0-4 as defined in step 2). - 'h': (0-4 as defined in step 2). - 'i': (0-4 as defined in step 2). - 'j': (0-4 as defined in step 2).

OpenAI API Batch Modality

The final dataset used for annotation was large, and annotating it using the standard OpenAI API endpoints would have been too time-consuming. To speed up the process, we used the Batch API, which allows users to upload a large number of requests at once. OpenAI processes these requests within 24 hours, optimizing for times of lower traffic.

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

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

A.4 Requirements and Sources

Table A.1: *Data Sources*

Data	Source	Download Date	Availability
Twitter Data			
Model-tweets dataset	Twitter Historical APIv2	7^{th} - 8^{th} Jan. 2022	Cannot be shared
Analysis-tweets dataset	Twitter Historical APIv2	4^{th} - 7^{th} Dec. 2022	Cannot be shared
GIS Data			
U.S. Shapefile (v4.1)	GADM website	4^{th} Oct. 2022	Can be shared
Election Data			
Presidential election dataset	FiveThirtyEight repository	17^{th} Nov. 2022	Can be shared

Notes: The table reports a description of the sources of the data used in our analysis and their respective availability. In Section 2.3 we describe the process of data preparation.

B Comparison of Narratives' Classification: GPT vs Human Coders

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

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

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

Method

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

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

The classification task was divided into two phases, matching as closely as possible the language and structure of the prompt used for GPT's task. For each character, the Human Coder was first asked whether the character was present in the tweet. For characters identified as present, coders

then indicated whether the character was depicted as a hero, villain, victim, or none of these roles. Finally, we used these classifications to compare Human Coders to each other and to GPT. Below, we present the results.

Comparison at the Character Level

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

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

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

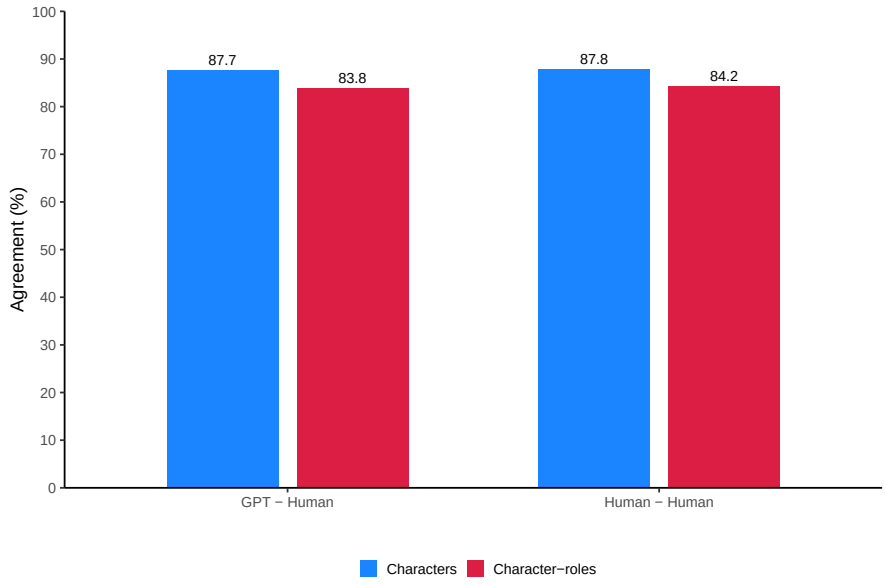
As explained above, our main agreement measure includes both positive and negative classifications. In most cases, characters are **not** present, so including negative agreement (where coders agree a character is absent) can inflate the overall agreement rates. To ensure the results are not driven by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

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

analysis (Figure B.1), agreement captured both uncertainty about the presence of characters and the assignment of roles. In other words, coders were compared on both whether a character was present and which role, if any, was assigned. In the next step, we select only those tweets where the two Human Coders agreed that a specific character was present, to then compare the agreement on the assignment of roles, both between the Human Coders and between GPT and Human Coders.

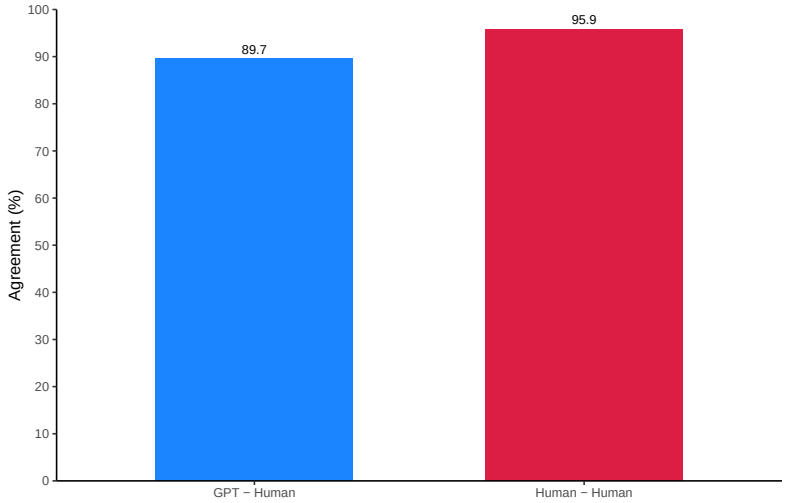
Figure B.2 shows the results of this second exercise. On the left, the blue bar indicates agreement between GPT and Human Coders; on the right, the red bar shows agreement between pairs of Human Coders, limited to tweets where both Human Coders agreed on the presence of the character. As expected, agreement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

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



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

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



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

Comparison at the Tweet Level

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

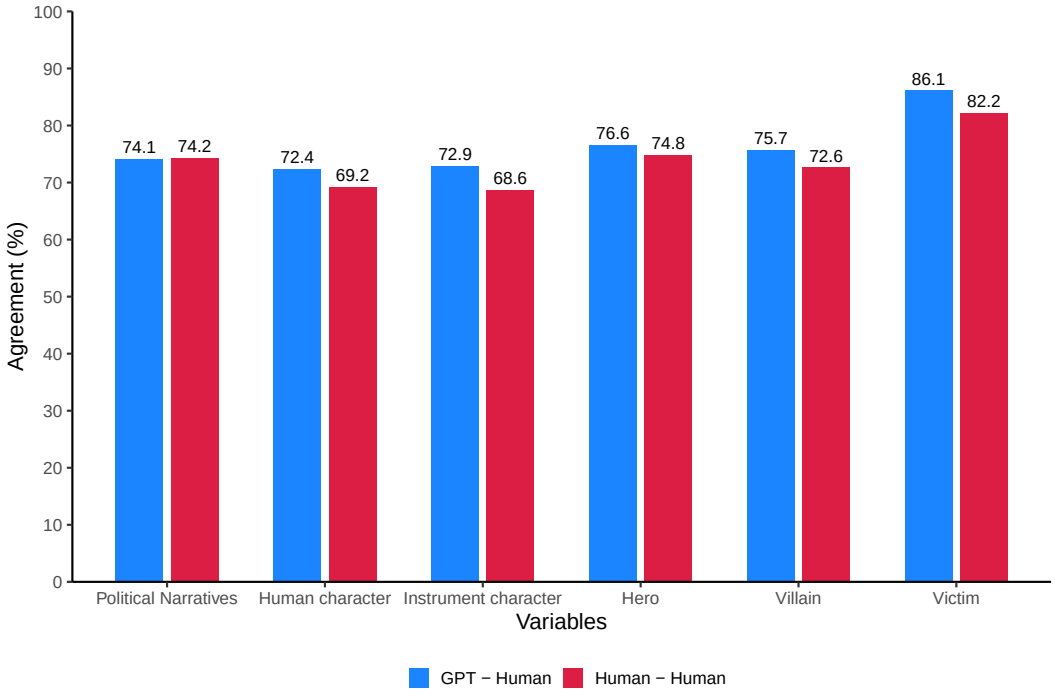
Political Narratives We compute the political narratives variable at the tweet level, defined as containing at least one character-role combination. We assess agreement rates between GPT and Human Coders (blue) and between pairs of Human Coders (red). As shown in Figure B.3, agreement levels are very similar: around 74% for GPT–Human Coder comparisons and 75% for Human Coder pairs. As above, we also compute Fleiss' κ to provide an unbiased measure. The κ value is 0.215, indicating fair agreement.

Human and Instrument Characters We use the classified tweets to compute two additional variables also used in our analysis: the Human Character variable, defined as containing at least one human character, and the Instrument Character variable, defined as containing at least one instrument character. We observe a high level of agreement in detecting both human and instrument characters. As shown in Figure B.3, on average, GPT and Human Coders agreed on the presence of

at least one human character in approximately 73% of tweets. Similarly, the inter-human agreement on detecting both human and instrument characters is about 69%. Notably, two Human Coders agree, on average, to a lower degree than when pairing GPT with any Human Coder. Fleiss' κ in this case is higher, at 0.43, 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 GPT and Human Coders agreed on the presence of heroes in roughly 76% of cases, villains in 76%, and victims in about 86%. Once again, Human Coders agreed with GPT more often than they agreed with each other. Fleiss' κ values are 0.508 for heroes, 0.494 for villains, and 0.266 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 GPT and Human Coders, and between pairs of Human Coders, for the classification of 500 randomly selected tweets. All tweets were classified by GPT as containing at least one of characters of interest in our study. Twenty-eight Human Coders, each coding a different number of tweets, classified the sample. Each tweet was coded by two Human Coders. Agreement is defined as the number of identical classifications over the number of total tweets. *Political Narratives* is defined as the level of agreement on the presence of at least one character-role in a tweet (Hero, Villain, Victim). *Human character* is defined as agreement on the presence of at least one human character. *Instrument character* is defined as agreement on the presence of at least one instrument character. *Hero*, *Villain*, *Victim* measures agreement for the presence of at least one character-role in the tweet. The blue bars indicate the average level of agreement between GPT and the two humans. The red bar indicates the average level of agreement between two humans.

C Additional Output: Observational Data

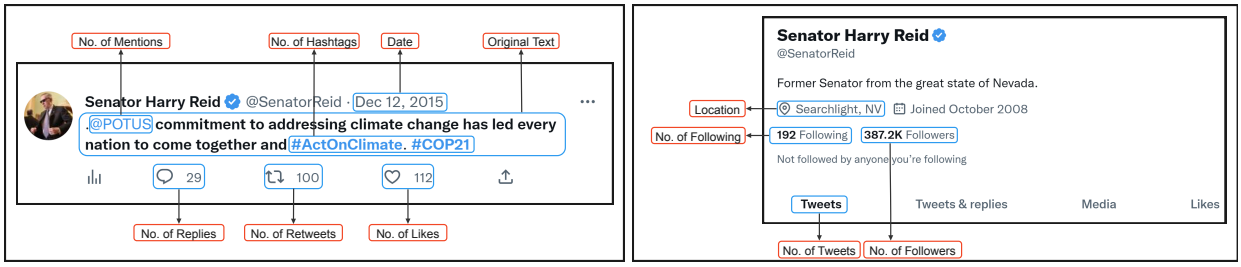
C.1 Additional Details on Variables

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

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

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

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



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

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

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

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

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

Variable	Question/Description	Categories/Scale/Interval	Source
Quality of Text			
Lexical Density	Ratio of content words (nouns, verbs, adjectives, adverbs) to total words	$\in [0.01; 0.92]$	Own computation
Type/Token Ratio	Ratio of unique words to total words in the tweet	$\in [0; 1]$	Own computation
Reading Ease	Flesch Reading Ease measure (higher = easier to read)	$\in [1; 114.63]$	Own computation
Education needed to comprehend text	Approx. U.S. 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 (U.S.), city}	Own computation

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

C.2 Additional Descriptive Statistics

In this section of Appendix C, we provide additional descriptive statistics on the observational data used to create the outputs in the paper and appendices. We begin by complementing the descriptive statistics on tweets shown in Table C.3 (and commented in the paper) with descriptive statistics on the users who posted those tweets, reported in Table C.4. For each variable, we report the mean, median, standard deviation, minimum, and maximum value, covering the level of localization, public metrics from the user's profile, and information from the profile description.

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

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

Notes: The table displays descriptive statistics for the dataset of relevant tweets used in our analysis. We define a tweet as relevant if it features at least one character from our list. We include only character roles that appear at least 100 times, thus excluding 'US Republicans-victim', 'Emission Pricing-victim', 'Regulations-victim', and 'Green Tech-victim'. For each variable, we report the average, median, standard deviation, and minimum/maximum values. We calculate the number of words per tweet excluding hashtags and mentions. We group variables by their role in the analysis. The appendix Table C.4 provides an overview of the users' characteristics, for those users that posted relevant tweets. The appendix Table C.5 shows the same statistics including all tweets that were classified by the GPT model, thus including also non-relevant tweets.

In Table C.4, 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. This is the most commonly used subset in our analysis, as we implement state fixed effects in most specifications and exclude users who could only be localized generically in the United States. 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 8,000 tweets (this refers to total activity since account creation, not within our dataset). In 14% of cases, users had a profile description and mentioned 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. Overall, our sample of users is skewed toward Democrats and individuals with higher education. However, this may reflect Republicans indicating their political leaning less often in their profile descriptions. This is the only information on which we base our measures of political leaning,

religiosity, education, and parenthood.

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

	Mean	Median	St. Dev.	Min.	Max.
Localization Level					
Share State-Located	.93	1	.25	0	1
Profile's Characteristics					
Share verified	.034	0	.18	0	1
No. of Followers	8,809	434	425,075	0	133,243,611
No. of Following	1,702	668	6,385	0	841,864
No. of Tweets	25,183	8,093	58,136	1	3,671,801
Profile Description					
Share of Democrats	.14	0	.35	0	1
Share of Republicans	.034	0	.18	0	1
Share religious	.056	0	.23	0	1
Share with High Educ.	.25	0	.43	0	1
Share with Children	.11	0	.31	0	1
No. of Observations	152,560				

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

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

Comparing the paper Table C.4 and Table C.5 some noteworthy points emerge. The subset of relevant tweets, compared to the totality of tweets, is generally more viral with retweets and like being almost twice as high. While relevant tweets are generally longer, the amount of hashtags and mentions used is comparable. Virtually all measures of text quality, emotions, and valence in text are comparable between the two datasets. Overall, the relevant tweets tend to be longer and more viral.

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

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

Notes: The table reports descriptive statistics for the dataset of all tweets used in the GPT classification. For each variable, we report the average, median, standard deviation, and minimum/maximum values. We 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'. We calculate the number of words per tweet excluding hashtags and mentions. We group variables by their role in the analysis.

C.3 Distribution of Narratives and Virality Outcomes

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

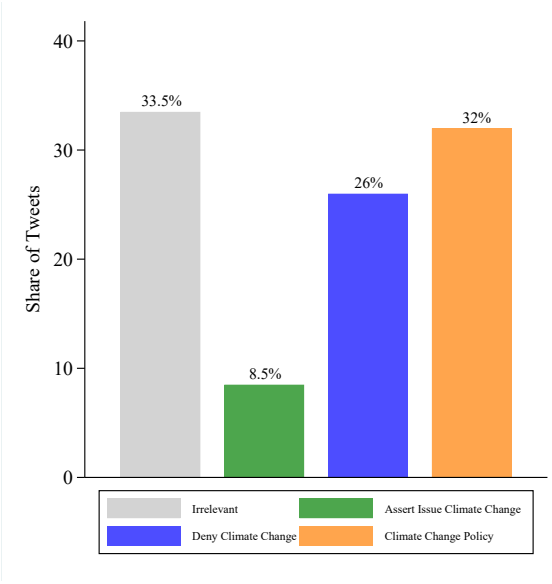
Figure C.2a shows the distribution of tweets classified in the first step of our pipeline. Roughly 33% of tweets are irrelevant, meaning they are either too unclear to classify, too short, or unrelated to climate change despite mentioning keywords from our list. Tweets classified as discussing climate change policy make up about 32% of all tweets. A large share of the conversation on Twitter/X revolves around simply asserting or denying the existence of man-made climate change. These tweets do not contribute to the policy debate but instead reflect fixed and polarized positions on

either side of the broader discussion.

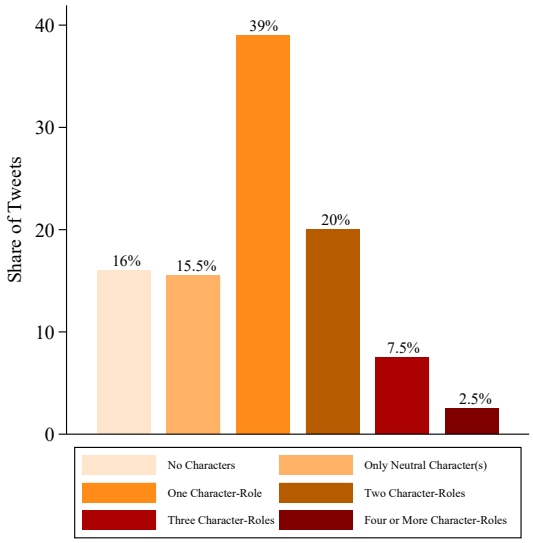
Only for tweets in the *Climate Change Policy* category do we prompt GPT to identify characters of interest. The distribution of these classifications is captured in Figure C.2a, where a clear picture emerges. The list of characters used in this analysis — developed through extensive reasoning and review of the literature and public debate — captures the discussion on climate change policy well. Our characters appear in the vast majority of tweets, with only 16% not including any of them. In 15.5% of tweets, characters are present but only in neutral form, meaning they are not assigned one of the three roles from the drama triangle. This subset serves as the comparison group in most of our analysis and forms the basis for estimating the effect of Political Narratives. About 39% of tweets contain at least one character-role, framing a character as a hero, villain, or victim. Two character-roles appear together in 20% of tweets, three in 7.5%, and four or more in 2.5%. Overall, the characters of interest appear in most policy-relevant tweets, with single character-roles being the most common.

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: On the right, Figure C.2b provides information on tweets that were categorized as being about climate policy. This includes tweets that are relevant, used in our analysis, and tweets that were classified as being about climate change policy discussion, but that do not contain any of our characters. We define a tweet as relevant if it features at least one character from our list. In the analysis, 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', indicated by a dot in the tables.

In the second part of this section, we provide additional details into the distribution of our character-roles. In other words, we dive into the classification of characters, divided into roles and neutral framing. We aim to complement the paper Table 2, which displays the distribution in shares, excluding those character-roles that did not reach at least 100 occurrences in the full set of classified tweets. We report the distributions in Table C.6.

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

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

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

Panel A: Human Characters

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

Panel B: Instrument Characters

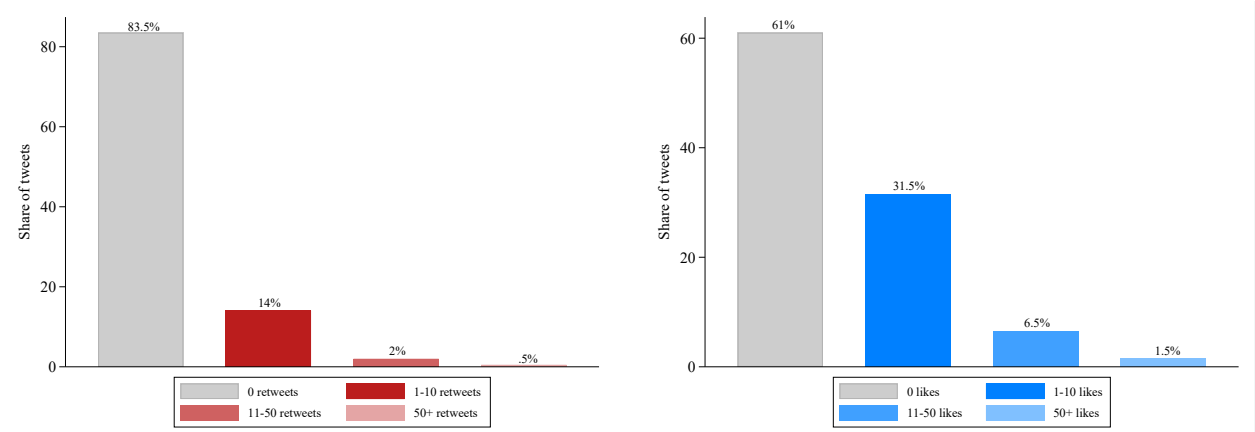
	Hero	Villain	Victim	Neutral	Total
Emission Pricing	11,955	7,337	11	6,985	26,288
Regulations	20,420	7,968	29	18,656	47,073
Fossil Industry	366	69,009	238	9,423	79,036
Green Tech	67,621	4,911	88	18,737	91,357
Nuclear Tech	5,077	1,879	107	2,577	9,640

Notes: The table displays the absolute frequencies of character-roles in the classified data. Sums are computed using the dataset of relevant tweets, used in our analysis. We define a tweet as relevant if it features at least one character from our list. In the analysis, 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', indicated by a dot in the tables. Panel (a) displays the sums for characters of the human type, while Panel (b) displays the same for characters of the instrument type. The column Neutral in both panels reports cases where the character is present in the tweet but is not depicted in one of the three specific roles. The occurrence of character-roles is not mutually exclusive, meaning multiple roles may appear in the same tweet. The main paper Table 2 displays similar information provide insights into shares, compute excluding the categories that do not reach 100 instances.

Figure C.3: ***Distribution of Retweets and Likes in Relevant Tweets (United States, 2010-2021)***

(a) Neutral Tweets

(b) Narrative Tweets



Notes: The figure provides insights into the nature of virality in the tweets from our sample. Figures C.3a and C.3b display the distribution of retweets across categories: 0, between 1 and 10, between 11 and 50, and 51 or more, respectively for neutral and for narrative tweets. For this analysis, we only include character-roles that appeared at least 100 times, thus excluding 'US Republicans-victim', 'Emission Pricing-victim', 'Regulations-victim' and 'Green Tech-victim'.

D Heterogeneous Effects

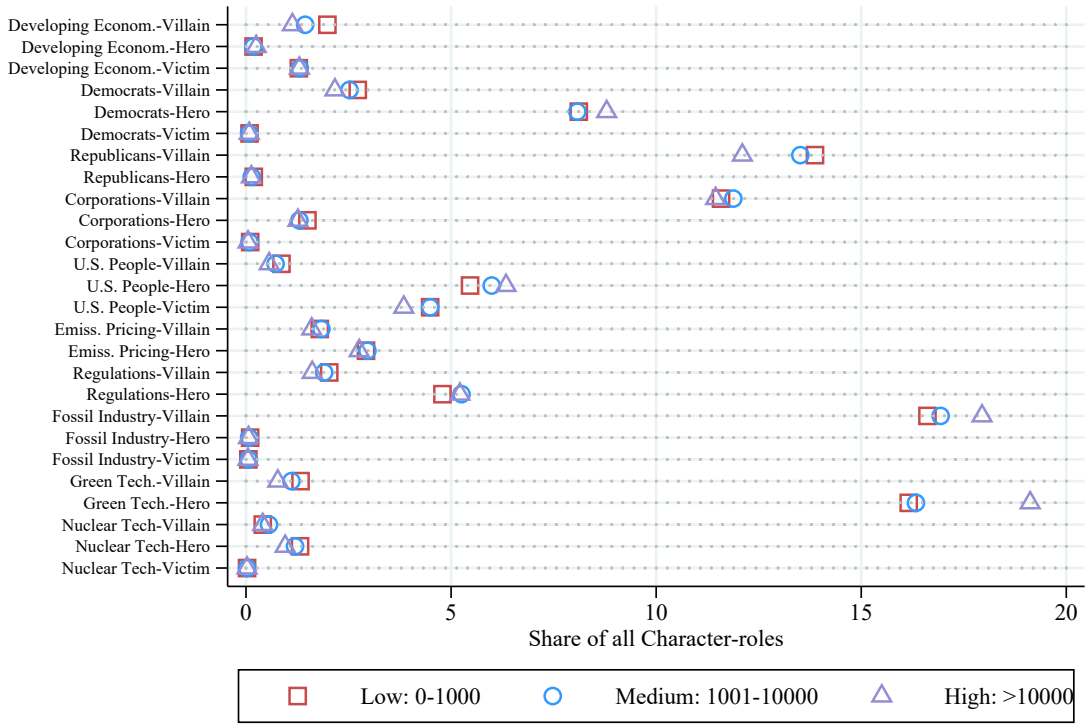
D.1 Heterogeneity by Visibility of the Profile

In this study we explore which narratives go viral and what are the factors determining this virality. In the section of the papers we address this question mostly approaching it from the tweet side. In other words we explore the impact of the content of tweet on the determination of the virality of these tweets. In this section of Appendix D we approach the issue from the side of the users. How does the users' visibility and influence within the platform impact the virality of their tweets?

We measure visibility by the number of followers on each user's profile. We divide users into three groups: low visibility (0 to 1,000 followers), medium visibility (1,001 to 10,000 followers), and high visibility (more than 10,000 followers). As a first step, we ask whether the types of narratives differ across users with different visibility levels. This check ensures that narrative content is comparable across visibility categories. Figure D.1 plots the share of each character-role relative to all character-roles within each visibility group. Squares represent low visibility profiles, circles medium visibility, and triangles high visibility. We plot the shares for all character-roles that appear at least 100 times in the full dataset.

Analyzing Figure D.1, it appears clear that there are virtually no differences in narrative content across profile categories, with only a few exceptions. Low and medium visibility profiles show extremely similar patterns. High visibility profiles differ in four cases: they post a slightly higher share of US Democrats–Hero narratives, a considerably lower share of US Republicans–Villain narratives, a higher share of Fossil Industry–Villain narratives, and a considerably higher share of Green Tech–Hero narratives. The latter is also the most widespread narrative used by high visibility users. Overall, differences across categories are rare, except that high visibility users appear slightly less politicized and more concerned with energy sources. A question remains: Does this lack of difference also translate in a similar impact of featuring narratives on the virality of their tweets?

Figure D.1: ***Share of Character-Roles by Visibility of the Profile***



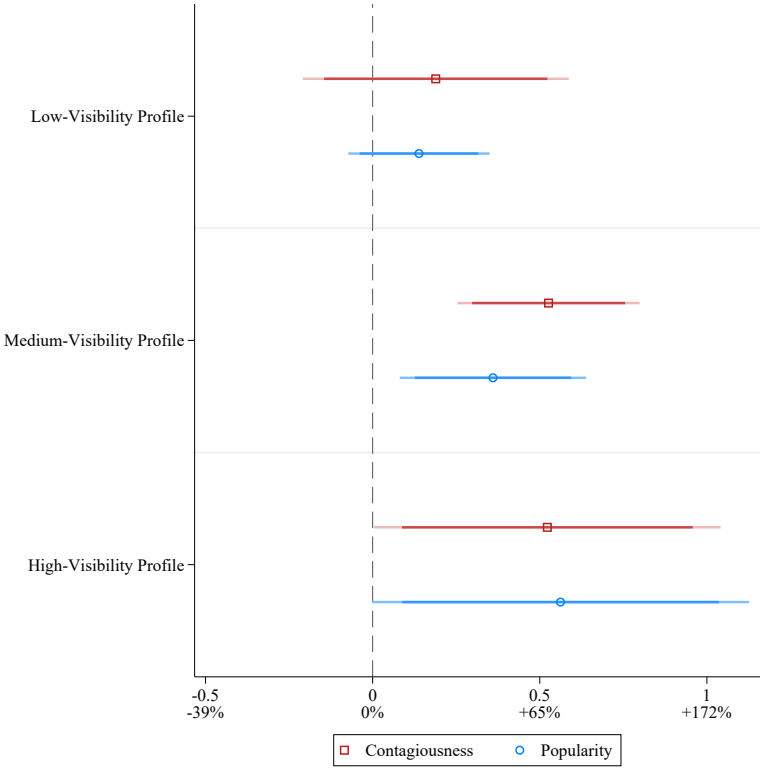
Notes: The figure shows the share of character-roles used by visibility of profiles. The square indicate shares for profiles with less than 1,000 followers, the circle indicates shares by profiles with number of followers between 1,001 and 10,000, and the triangle indicates the share for profiles with more than 10,000 followers. Shares are computed by splitting the sample into the desired profile categories, counting the number of time each character-role is present in their tweet, and dividing by the total amount of character-roles used by that category of profile. The descriptive statistics are computed using 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'.

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

This exercise provides some interesting results. Despite larger confidence intervals for the high visibility group — due to a smaller number of users — medium and high visibility profiles show

similar patterns. In both groups, narratives increase virality in line with the main results of the paper. In contrast, tweets from low visibility profiles show no measurable impact of narratives on virality. This finding lends itself to several interpretations. The most likely is that at low levels of visibility, the content posted by these users rarely goes viral, regardless of whether it contains a narrative or presents neutral framing. In other words, at low follower counts, factors other than content — such as limited reach or engagement — may prevent virality, making the narrative effect negligible.

Figure D.2: ***Regression Results - Impact of Political Narratives on Virality by Visibility of Profile***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood 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, and likes, Popularity. The sample of users is split into three groups, and the results are shown in the three panels of the figure. On top, results for users with low reach, namely number of follower below 1,000, the second panel for medium reach profiles, with number of followers between 1,001 and 10,000, and the bottom panel for high reach profile with more than 10,000 followers. 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 along with the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. All regressions include character fixed effects. We include hour, week and year-state fixed effects. Standard errors are clustered at the week level.

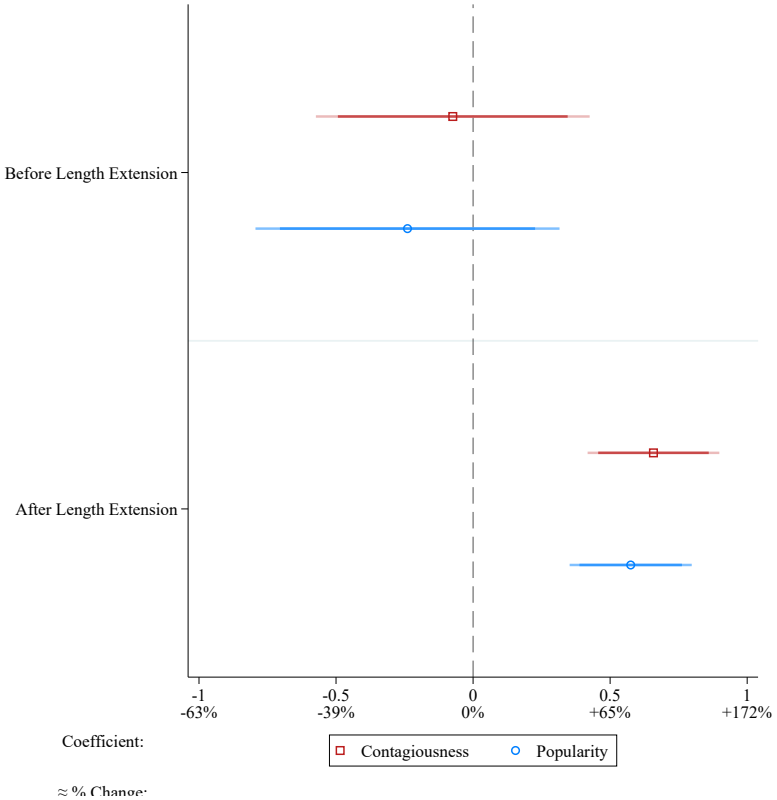
A note of caution is necessary for this analysis. The information on followers count reflects the moment of data extraction, which means it is an ex-post measure relative to when the tweets were posted. It is therefore possible that the visibility of some profiles results from the use of narratives,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

be interpreted carefully.

D.2 Heterogeneity by Length of Tweets

On 7^{th} November 2017 Twitter's policy on the allowed length of tweets changed. The policy changed allowed to go from a maximum of 140 character to a maximum of 280 character per tweet.

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



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood 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, and likes, as a robustness check. The sample is split into two groups, and the results are shown in the two panels of the figure. On top, results for the sample of tweets posted before November 7^{th} 2017, the panel on bottom shows effect for tweets posted after. The date coincides with the change in the maximum limit of characters per tweet from 140 to 280. 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 along with the corresponding percentage change rounded to the closest unit and computed as follows: $\approx e^{\beta} - 1$. All regressions include authors' characteristics (verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status), and character fixed effects. We include hour, week of the year and year-state fixed effects. Standard errors are clustered at the week level.

E Robustness Checks: Observational Data

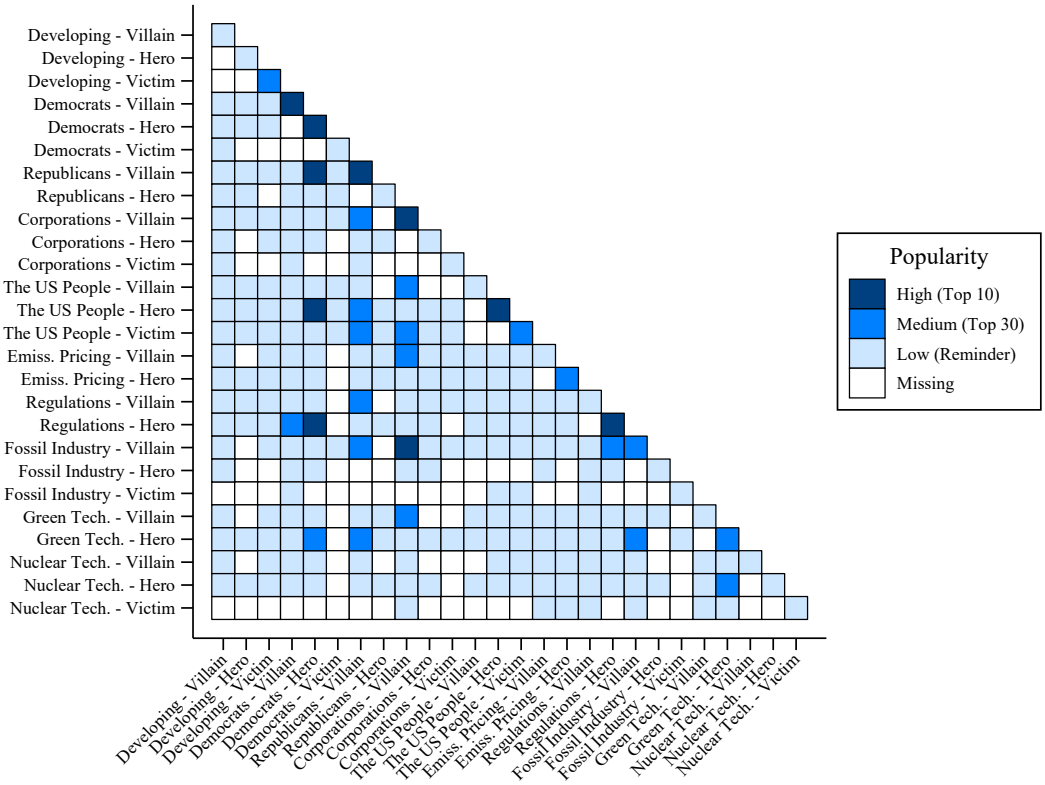
E.1 Impact of Narratives on Popularity

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

We begin our analysis on Popularity by replicating the same descriptive exercise from Section 4.2, now using the rate of likes instead of retweets. Figure E.1 shows a heatmap of like rates for all narratives that include one or two character-roles. Each square in the matrix reports the rate of likes for a specific pair of character-roles — or for a single character-role when it appears alone, along the diagonal. The rate is computed by summing all likes received by tweets featuring that specific combination (or single character-role), and dividing by the total number of likes across all tweets with one or two character-roles. The matrix is symmetrical. To keep the figure readable, we do not display the numerical values directly; instead, we highlight the top 10 and top 30 most frequent combinations using a color scale.

Figure E.1 shows that Popularity rates follow patterns similar to those observed for Virality (retweets), though with some notable differences. While the most polarized and politicized narratives - those on Democrats and Republicans - still drive engagement, People-hero narratives stand out even more strongly in this dimension, ranking among the top five both independently and when combined with Democrats as heroes. Across both Virality (retweets) and Popularity (likes), human characters tend to be more viral, yet Regulations and Green Tech also feature prominently, appearing in several of the top 30 most popular narratives.

Figure E.1: ***Popularity (Likes) of Character-Roles - Tweets with One or Two Character-Roles***



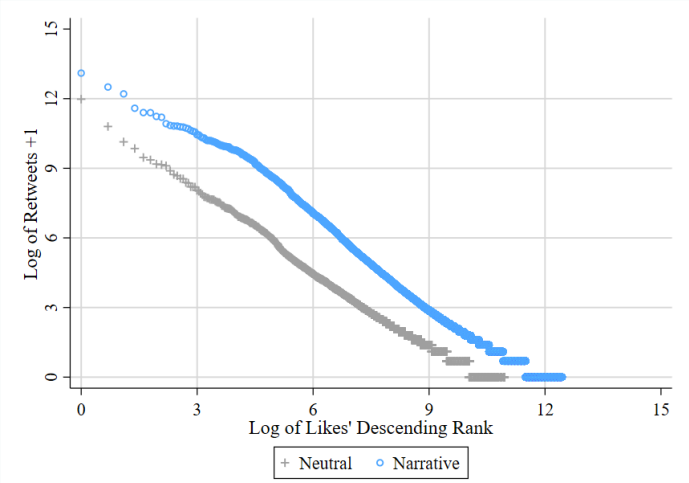
Notes: The figure shows the like rate of each character-role that appears alone or alongside another character-role in relevant tweets that contain one or two character-roles, which account for nearly 60% of policy-relevant tweets. We define a tweet as relevant if it features at least one character from 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'. Like rates are computed by dividing the total number of likes received by a particular pair or single character-role by the total number of likes of tweets containing one or two character-roles. The diagonal of the matrix shows the like rate when each character-role appears alone. Tweets with three or more character-roles are excluded. To avoid visual overload, we do not display exact rates. Instead, we use a color scheme to highlight the top 10 most frequently liked 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: 'Democrats-hero' (13.90%), 'Republicans-villain' (13.90%), 'U.S. People-hero' (7.00%), 'Republicans-villain + Democrats-hero' (6.13%), 'U.S. People-hero + Democrats-hero' (5.32%), 'Fossil Industry-villain + Corporations-villain' (5.20%), 'Democrats-villain' (3.85%), 'Regulations-hero + Democrats-hero' (3.21%), 'Regulations-hero' (2.29%), 'Corporations-villain' (2.18%)

Building on the analysis in the main paper, we take a step further and examine how narratives influence the distribution of likes. Specifically, we compare tweets that feature a Political Narrative with those that do not. Figure E.2 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 or retweeted tweet in the dataset. The y-axis reports the logarithm of the number of retweets or likes, plus one.

The figure shows that the distribution of Popularity also follows a Power Law, as in 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 to 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 sharper. Overall, this descriptive evidence indicates that Popularity follows similar patterns to Virality, and that tweets featuring narratives consistently attract more likes than neutral ones.

Figure E.2: ***Popularity of Political Narratives in Relevant Tweets (United States, 2010-2021)***

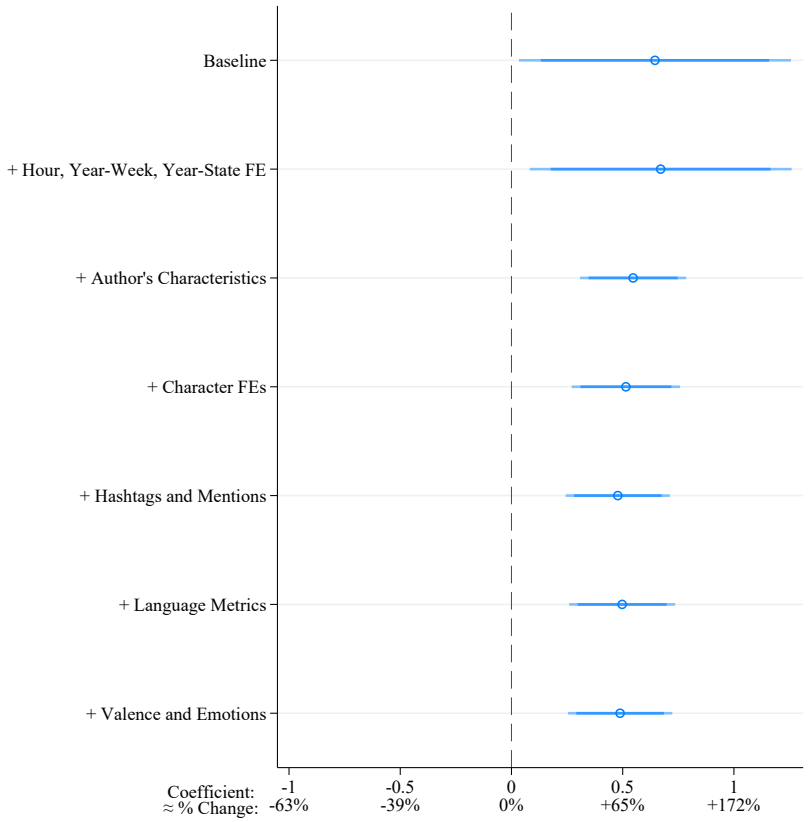


Notes: The figure provides insights into the nature of virality in the tweets from our sample. It displays the Log-Log Rank Distribution of likes for relevant tweets. We define a tweet as relevant if it features at least one character from our list. We include only character-roles that appear at least 100 times, thus excluding 'US Republicans-victim', 'Emission Pricing-victim', 'Regulations-victim', and 'Green Tech-victim'. The x-axis represents the logarithm of the rank distribution, where rank 1 corresponds to the most liked tweet in the sample. The y-axis represents the logarithm of the number of likes, plus one. The slope of the curve reflects how rapidly the dataset transitions from the tweets with most likes, to the least. The vertical position of the curve at a given rank provides a measure of relative virality—tweets with a higher curve position receive more engagement than those at the same rank in a dataset with a lower curve. The Paper Figure 5 shows the same for retweets.

We turn next to regression analysis, extending the descriptive findings by estimating the same models used in Figure 7, this time applied to Popularity. Figure E.3 presents a coefficient plot based on Poisson Pseudo-Maximum Likelihood estimates. The models follow a sequential design, becoming progressively more restrictive from the top to the bottom panel. Each panel builds on the previous one by adding further controls, as noted in the titles. As before, the coefficients can be interpreted as approximate percentage changes using the transformation: $\approx e^{\beta} - 1$.

The results align closely with those presented in the main paper for tweet Virality. Consistent with the findings on retweets, Political Narratives have a clear positive effect on Popularity. Across all model specifications, tweets that include at least one character-role receive more likes than those that present characters in a neutral way. As in the main analysis, user characteristics emerge as an important control: accounting for them sharpens the results and helps reduce noise in the estimation.

Figure E.3: ***Regression Results - Impact of Political Narratives on Conversation***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regressions testing the effect of featuring at least one character-role, compared to featuring characters only in a neutral role, on conversation, measured as the count of replies a post gets. 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 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 baseline 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 model controls for tweet characteristics (number of hashtags and mentions). The sixth and seventh models include, respectively, language metrics and valence/emotions, as used in the descriptive analysis above. The Paper Figure 7 shows the same for Virality (retweets) and (likes) as a robustness check.

In the final part of this robustness check, we take a closer look at Political Narratives by examining the individual impact of the three drama triangle roles: Hero, Villain, and Victim. Table E.1 reports Poisson Pseudo-Maximum Likelihood estimates of how each role affects the number of likes received by tweets. The analysis is restricted to tweets where roles are mutually exclusive — that is, if a tweet features a Hero, it does not also include a Villain or a Victim. Columns 1 to 3 compare Hero, Villain, and Victim narratives separately to tweets with only neutral framing. Columns 4 to 7 include all roles simultaneously, changing the reference category in each column: first neutral (column 4), then Hero, Villain, and Victim, respectively.

When comparing these results to those found for Virality, several meaningful patterns emerge, highlighting both parallels and distinctions between the effects on retweets and likes. In both cases, Hero and Villain narratives significantly boost engagement relative to tweets that feature characters in neutral roles. Column 1 of Table E.1 shows that Hero narratives increase the number of likes by approximately 53% (computed as $\approx e^{\beta} - 1$ for Poisson models). Villain narratives, however, produce an even stronger effect, with a coefficient nearly twice as large. These effect sizes closely match those observed for retweets, indicating that narratives centered around Heroes and Villains consistently draw higher engagement. By contrast, Victim narratives do not show a statistically significant effect on either metric, suggesting that this role is less effective at driving user interaction.

When looking at the full models in columns 4 through 7, a more nuanced picture emerges. Villain narratives consistently stand out, outperforming both Hero and Victim narratives in driving engagement. This supports the idea that narratives built around conflict and opposition tend to be the most effective at capturing attention. The effect of Hero narratives remains statistically significant only when compared to neutral tweets. This suggests that while Heroes can boost engagement, their impact is less pronounced when set against the stronger pull of Villain narratives.

Lastly, we reproduce the results on the interactions of roles in Figure E.4. Each panel presents the results of a single Poisson Pseudo-Maximum Likelihood regression model. In each model, the comparison group consists of tweets featuring only one role and only one character-role (for example, a single hero narrative). The regressors capture all combinations where that role appears alongside other roles (e.g., hero + villain, hero + victim, hero + villain + victim) or where the same role applies to multiple characters (e.g., two different hero character-roles within the same tweet).

The results for Virality (retweets) and Popularity (likes) as a robustness check are strikingly similar. Compared to tweets featuring a single-character hero narrative, increasing complexity by introducing a hero-victim or hero-victim-villain structure reduces overall virality. However, when the hero role is combined exclusively with a villain, virality increases, while popularity remains unchanged. A similar pattern emerges for villain narratives, where complexity tends to reduce virality, particularly when the full Drama Triangle is present. Notably, when villains are reinforced within a simple narrative —meaning multiple characters share the villain role— virality increases even further, suggesting that narratives centered around multiple antagonists tend to gain more traction. For victim narratives, no consistent pattern emerges, except when a victim is paired with a villain, which significantly boosts virality.

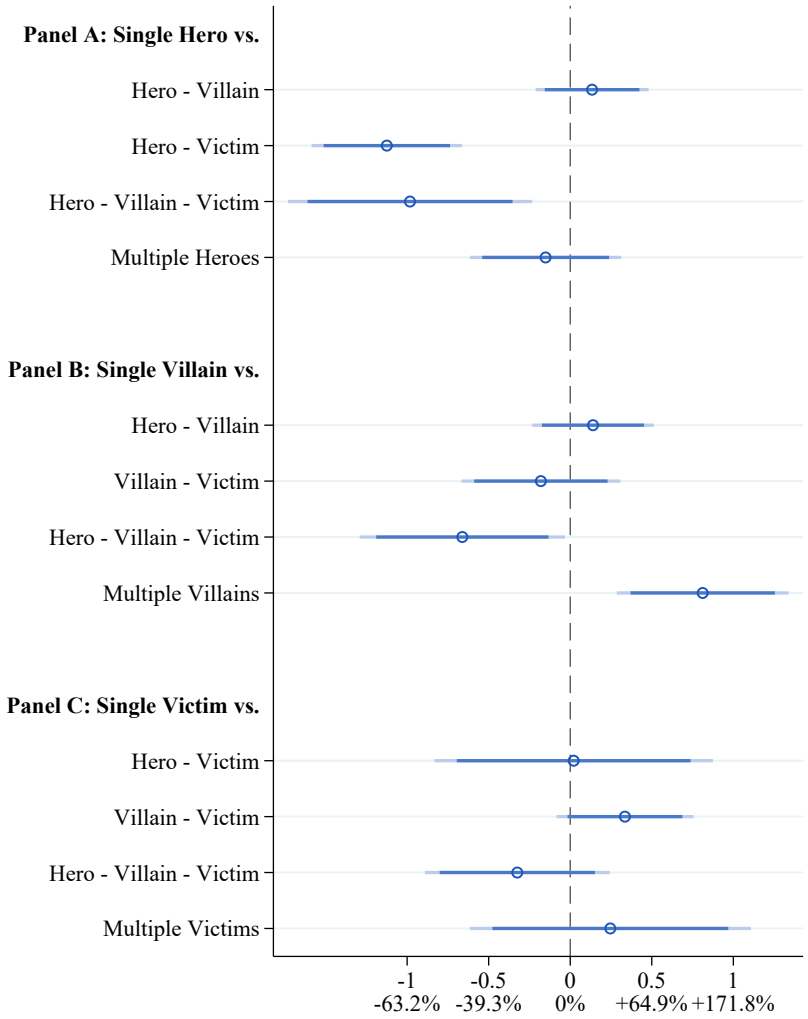
Overall, the results on conversation are strongly consistent with those for Virality. Villain narratives appear even more important in driving virality through conversation. However, featuring multiple victims also has a positive effect on the number of comments.

Table E.1: ***Regression Results - Impact of Individual Roles on Popularity***

Dependent Variable	**Replies' Count**						
	Coeff./SE/p-value						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hero	0.430			0.434		-0.518	0.100
	(0.211)			(0.214)		(0.280)	(0.288)
	[0.042]			[0.042]		[0.064]	[0.728]
Villain		0.823		0.952	0.518		0.618
		(0.123)		(0.250)	(0.280)		(0.321)
		[0.000]		[0.000]	[0.064]		[0.054]
Victim			0.073	0.334	-0.100	-0.618	
			(0.306)	(0.267)	(0.288)	(0.321)	
			[0.811]	[0.210]	[0.728]	[0.054]	
Neutral					-0.434	-0.952	-0.334
					(0.214)	(0.250)	(0.267)
					[0.042]	[0.000]	[0.210]
Sample: Neutral	✓	✓	✓	✓	✓	✓	✓
Sample: Hero	✓			✓	✓	✓	✓
Sample: Villain		✓		✓	✓	✓	✓
Sample: Victim			✓	✓	✓	✓	✓
Mean Outcome Reference Group	9.523	9.523	9.523	9.523	34.136	16.535	13.269
Pseudo R2	0.88	0.72	0.70	0.80	0.80	0.80	0.80
Obs	74739	88417	43084	134948	134948	134948	134948

Notes: The table show the coefficients of Poisson Pseudo-Maximum Likelihood regressions testing the effect of featuring Hero, Villain, and Victim roles on Popularity, measured as the count of likes. The regression models include relevant tweets featuring at most one character, either neutral or in a role. 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'. The regressions' coefficients can be transformed to percentage change with the formula: $\approx e^{\beta} - 1$. All regressions include authors' characteristics (verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status), and character fixed effects. We include hour, week of the year and year-state fixed effects. Standard errors are clustered at the week level, covering all weeks in our time frame.

Figure E.4: ***Regression Results - Heterogeneity in the Impact of Character-Role Combinations on Conversation***



Notes: The figure shows the coefficients from Poisson Pseudo-Maximum Likelihood regressions testing the effect of character-role combinations on Popularity, measured as number of likes. 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 dependent variable is the tweet-level count of retweets. The x-axis reports coefficient estimates along with the corresponding approximate percentage change, computed as follows: $\approx e^{\beta} - 1$. In both figures, each panel corresponds to a regression analyzing the effects of specific roles. Panel A examines the effects for hero, with the comparison group being tweets featuring a single hero character. The regressors capture character-role combinations where the hero is paired with others. Specifically, we estimate the effects of a hero paired with a villain or more villains, a hero paired with a victim or more victims, a hero paired with both villains and a victims, and tweets featuring multiple heroes. Panels B and C follow the same structure, focusing on villain and victim, respectively, with their corresponding comparison groups. All models include hour, week of the year, and year-state fixed effects. Standard errors are clustered at the week level, covering all weeks in the time period. The Paper Figure 8 show the same for Virality (retweets).

E.2 Impact of Narratives on Conversation

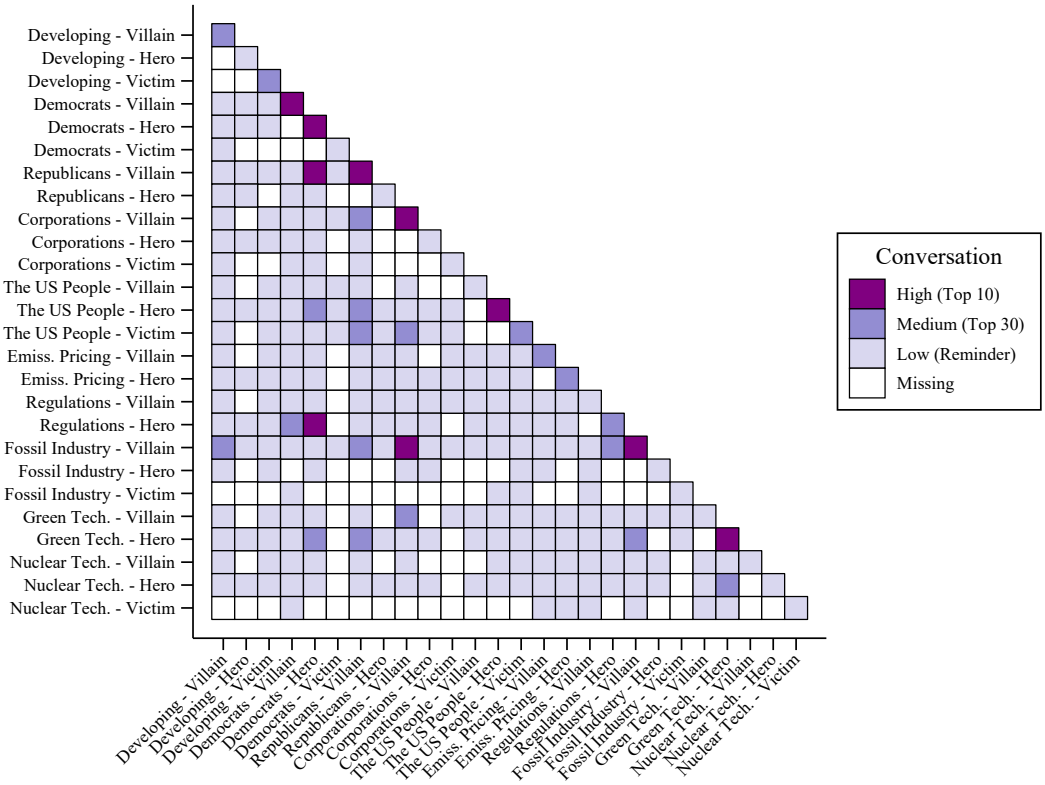
In this section of Appendix E.2, we provide a robustness check on the impact of narratives by examining their effect on conversation. We measure conversation using the count of comments a post receives. We do not include this outcome in the main analysis, as it captures a slightly different concept than virality. While a high number of comments may reflect popularity and engagement, it can also result from a small number of users exchanging opinions or arguing in the comment section. Despite being a less precise measure of virality, conversation provides a valuable robustness check.

We begin this exercise by exploring descriptive statistics on the rate of conversation for narratives featuring one or at most two character-roles, in line with the main results of the paper. Figure E.5 mirrors its correspondents for Virality and Popularity (likes) as a robustness check presented in Section 4.2. It shows a heatmap of conversation rates for all combinations of one or two character-roles. The color of each square reflects the conversation rate for that combination, calculated by dividing the total number of comments received by tweets with that combination by the total number of comments across all tweets featuring one or two character-roles. The diagonal indicates the conversation rate for each character-role occurring alone. The matrices are symmetrical. Numerical values are not displayed to avoid visual overload, but the top 10 and top 30 most frequent combinations are highlighted using a color scheme.

Figure E.5 suggests a pattern in conversation rates that closely mirrors the patterns observed for Virality and the robustness specification Popularity (likes). As with retweets and likes, the highest rates of conversation cluster around the more politicized narratives featuring Democrats and Republicans. High conversation rates also appear around some of the instrument characters. Regulations, the fossil industry, and Green Tech prompt many comments, especially in tweets framing the fossil industry as the Villain of the climate change debate.

One noticeable difference from the patterns observed for Virality is the prominence of Emission Pricing in conversation. Whether presented as Hero or Villain, Emission Pricing tends to spark more discussion than it generates retweets or likes. This suggests that when the debate shifts toward policy instruments — such as pricing emissions — users engage in more elaborate discussions rather than simply expressing support for a polarized and political side. However the difference is less clear with Popularity, where Emission Pricing-hero is also in the top 30 for rate of likes.

Figure E.5: ***Conversation (Replies) on Character-Roles - Tweets with One or Two Character-Roles***



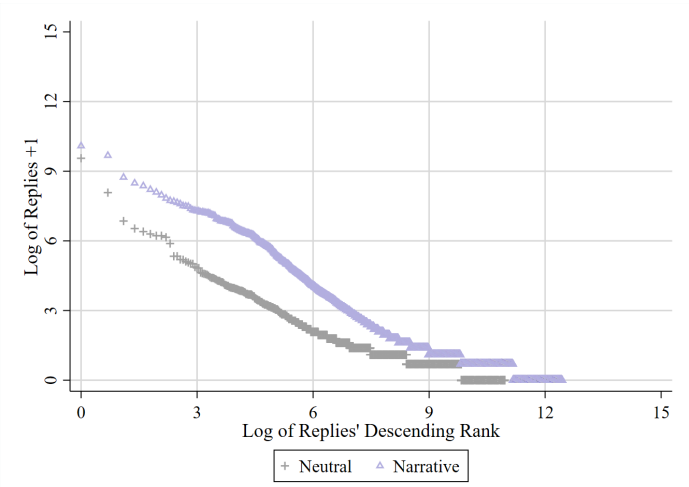
Notes: The figure shows the reply rate of each character-role that appears alone or alongside another character-role in relevant tweets that contain one or two character-roles, which account for nearly 60% of policy-relevant tweets. We define a tweet as relevant if it features at least one character 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'. Reply rates are computed by dividing the total number of replies received by a particular pair or single character-role by the total number of replies of tweets containing one or two character-roles. The diagonal of the matrix shows the reply rate when each character-role appears alone. Tweets with three or more character-roles are excluded. To avoid visual overload, we do not display exact rates. Instead, we use a color scheme to highlight the top 10 most frequently liked 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: 'Democrats-hero' (22.33%), 'Republicans-villain' (11.64%), 'Democrats-hero + Republicans-villain' (4.99%), 'U.S. People-hero' (6.24%), 'Democrats-villain' (4.45%), 'Fossil Industry-villain' (4.07%), 'Green Tech-hero' (4.03%), 'Corporations-villain' (3.01%), 'Fossil Industry-villain + Corporations-villain' (2.85%), 'Regulations-hero + Democrats-hero' (2.64%).

We take a step forward into the analysis of the impact of narratives on conversation by exploring the distribution of comments in tweets featuring a narrative and tweets featuring characters only in neutral framing. Similar to what is presented in Section 4.2 of the paper for Virality (retweets) and the robustness check Popularity (likes) , Figure E.6 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 follow a power law, consistent with what we observe for retweets and likes. 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 and likes, the curves are flatter and shifted downward, reflecting the smaller number of comments overall. Lastly, consistent with the results on Virality (retweets)and Popularity (likes) as a robustness check, the figure provides an initial indication of the effect of narratives. The curve for narrative tweets lies consistently above that for neutral tweets. At each rank, tweets featuring a narrative receive more comments, suggesting that Political Narratives may not only spark more support in the form of likes and retweets but also encourage greater engagement through conversation.

Figure E.6: ***Conversation on Political Narratives in Relevant Tweets (United States, 2010-2021)***



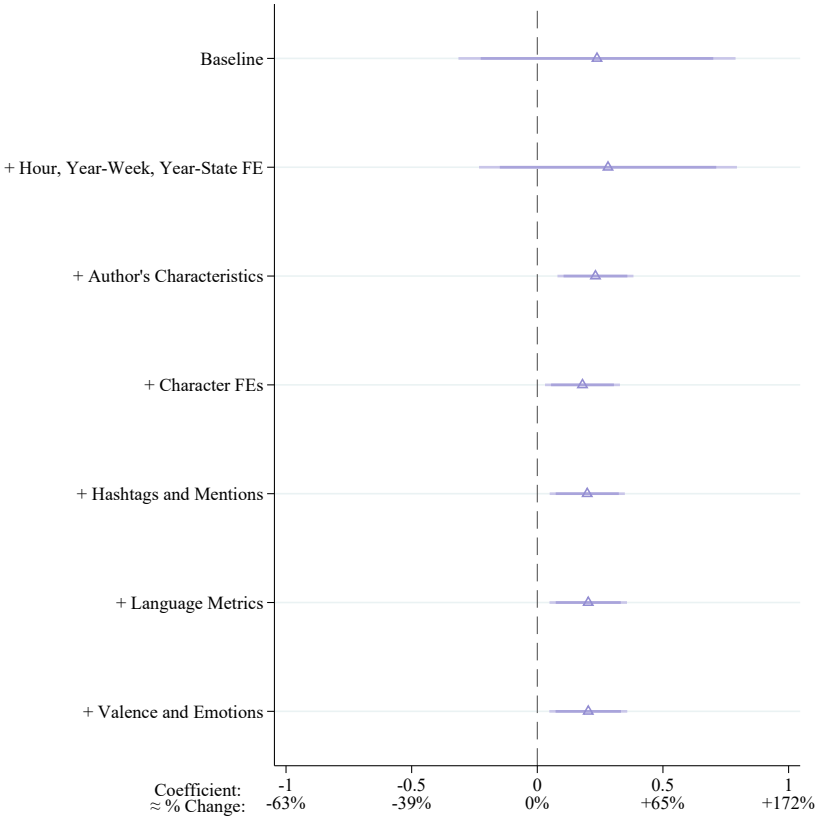
Notes: The figure provides insights into the nature of virality in the tweets from our sample. It displays the Log-Log Rank Distribution of replies for relevant tweets. We define a tweet as relevant if it features at least one character from our list. We include only character-roles that appear at least 100 times, thus excluding 'US Republicans-victim', 'Emission Pricing-victim', 'Regulations-victim', and 'Green Tech-victim'. The x-axis represents the logarithm of the rank distribution, where rank 1 corresponds to the most retweeted tweet in the sample. The y-axis represents the logarithm of the number of replies, plus one. The slope of the curve reflects how rapidly the dataset transitions from the tweets with most replies, to the least. The vertical position of the curve at a given rank provides a measure of relative virality—tweets with a higher curve position receive more engagement than those at the same rank in a dataset with a lower curve.

We move from descriptive to regression analysis by reproducing the results of Figure 7 for conversation. Figure E.7 presents a coefficient plot showing the results of models that mirror those used for Contagiousness (retweets) and Popularity (likes) as a robustness check. The models are estimated using Poisson Pseudo-Maximum Likelihood and apply 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 (retweets) and Popularity (likes) as a robustness check. 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 where 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 effect is about half that observed for retweets and likes, but narratives still appear to play an important role in driving conversation.

Figure E.7: ***Regression Results - Impact of Political Narratives on Conversation***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regressions testing the effect of featuring at least one character-role, compared to featuring characters only in a neutral role, on conversation, measured as the count of replies a post gets. 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 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 baseline 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 model controls for tweet characteristics (number of hashtags and mentions). The sixth and seventh models include, respectively, language metrics and valence/emotions, as used in the descriptive analysis above.

In the final part of this exercise, we use the flexibility of our framework to move beyond the overall effect of Political Narratives. We ask which roles drive the increase in conversation and how the drama triangle roles interact to shape this effect. Table E.2 presents Poisson Pseudo-Maximum Likelihood regression models estimating the impact of the drama triangle roles on the number of

comments received by tweets. All models include only tweets where roles are mutually exclusive, that is, e.g. if a hero appears in a tweet, that tweet does not contain a villain or a victim. The first three columns compare, respectively, hero, villain, and victim narratives to tweets featuring only neutral framing. Columns 4 to 7 include all roles at once, changing the comparison group for each column: starting with neutral (column 4), then hero, villain, and victim, respectively.

The results of these models provide a clear picture. Unlike what we observed for Virality (retweets) and Popularity (likes) as a robustness check, when focusing on mutually exclusive roles, only villain narratives have a meaningful impact on the amount of conversation. Whether directly compared to neutral tweets or in the full models, hero and victim narratives do not significantly increase the number of comments. The prominence of villain narratives may have several explanations, though it is important to remember that this is only correlational evidence. These results support the intuition proposed earlier: conversation may be driven more by argument and confrontation, which villain narratives are more likely to provoke, rather than by simple expressions of support or empathy.

Lastly, we reproduce the results on the interactions of roles in Figure E.8. Each panel collects the results of a single Poisson Pseudo-Maximum Likelihood regression model, where the comparison group correpsonds to tweet featuring only a single type of narratives e.g. hero narratives, and the regressors represent all possible compbinations of that specific type of narratives, e.g. hero plus villain, hero plus victim, hero plus villain and victim, and hero plus other hero narratives.

Lastly, we reproduce the results on the interactions of roles in Figure E.8. Each panel presents the results of a single Poisson Pseudo-Maximum Likelihood regression model. In each model, the comparison group consists of tweets featuring only one role and only one character-role (for example, a single hero narrative). The regressors capture all combinations where that role appears alongside other roles (e.g., hero + villain, hero + victim, hero + villain + victim) or where the same role applies to multiple characters (e.g., two different hero character-roles within the same tweet).

The models reveal patterns similar to those observed for Virality (retweets) and Popularity (likes) as a robustness check. Although not always statistically significant, the results suggest that adding complexity to narratives often reduces virality. For all three roles — hero, villain, and victim — when all roles are featured together, forming a more nuanced narrative, the number of comments decreases significantly. Despite smaller effect sizes, the impact of multiple villains aligns with the results for retweets and likes: more villains tend to spark more comments. The only exception to this pattern is the victim role. Unlike for retweets and likes, tweets featuring multiple victims seem to increase conversation compared to those with a single victim.

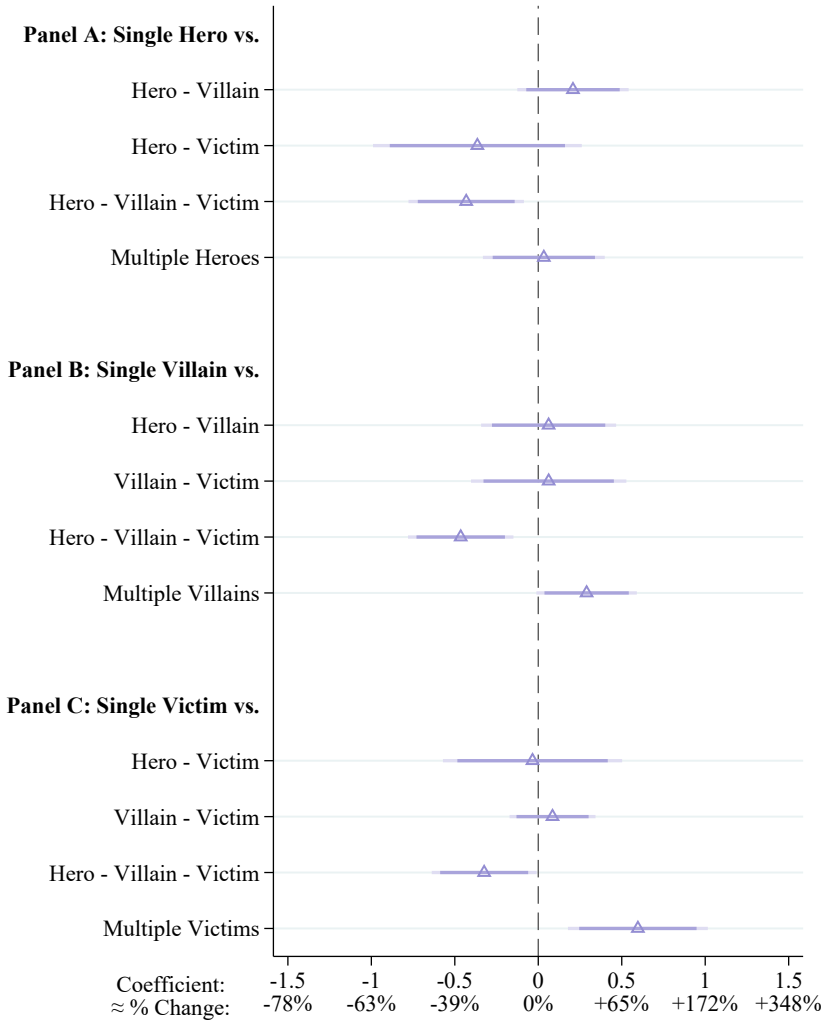
Overall, the results on conversation are strongly consistent with those for Virality (retweets) and Popularity (likes) as a robustness check. Villain narratives appear even more important in driving virality through conversation. However, featuring multiple victims also has a positive effect on the number of comments.

Table E.2: ***Regression Results - Impact of Individual Roles on Conversation***

Dependent Variable	**Replies' Count**						
	Coeff./SE/p-value						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hero	0.076			0.019		-0.352	-0.081
	(0.118)			(0.132)		(0.142)	(0.203)
	[0.521]			[0.887]		[0.013]	[0.690]
Villain		0.243		0.371	0.352		0.271
		(0.071)		(0.116)	(0.142)		(0.187)
		[0.001]		[0.001]	[0.013]		[0.147]
Victim			0.130	0.100	0.081	-0.271	
			(0.165)	(0.163)	(0.203)	(0.187)	
			[0.431]	[0.540]	[0.690]	[0.147]	
Neutral					-0.019	-0.371	-0.100
					(0.132)	(0.116)	(0.163)
					[0.887]	[0.001]	[0.540]
Sample: Neutral	✓	✓	✓	✓	✓	✓	✓
Sample: Hero	✓			✓	✓	✓	✓
Sample: Villain		✓		✓	✓	✓	✓
Sample: Victim			✓	✓	✓	✓	✓
Mean Outcome Reference Group	1.972	1.972	1.972	1.972	4.350	3.683	2.395
Pseudo R2	0.79	0.54	0.49	0.70	0.70	0.70	0.70
Obs	74554	88592	43191	134041	134041	134041	134041

Notes: The table show the coefficients of Poisson Pseudo-Maximum Likelihood regressions testing the effect of featuring Hero, Villain, and Victim roles on Conversation, measured as the count of replies. The regression models include relevant tweets featuring at most one character, either neutral or in a role. 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'. The regressions' coefficients can be transformed to percentage change with the formula: $\approx e^{\beta} - 1$. All regressions include authors' characteristics (verified status, number of followers and followings, total tweets created, party affiliation as Democrat or Republican, religiosity, higher education, and parenthood status), and character fixed effects. We include hour, week of the year and year-state fixed effects. Standard errors are clustered at the week level, covering all weeks in our time frame.

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



Notes: The figure shows the coefficients from Poisson Pseudo-Maximum Likelihood regressions testing the effect of character-role combinations on Conversation, measured as number of replies. 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 dependent variable is the tweet-level count of retweets. The x-axis reports coefficient estimates along with the corresponding approximate percentage change, computed as follows: $\approx e^{\beta} - 1$. In both figures, each panel corresponds to a regression analyzing the effects of specific roles. Panel A examines the effects for hero, with the comparison group being tweets featuring a single hero character. The regressors capture character-role combinations where the hero is paired with others. Specifically, we estimate the effects of a hero paired with a villain or more villains, a hero paired with a victim or more victims, a hero paired with both villains and a victims, and tweets featuring multiple heroes. Panels B and C follow the same structure, focusing on villain and victim, respectively, with their corresponding comparison groups. All models include hour, week of the year, and year-state fixed effects. Standard errors are clustered at the week level, covering all weeks in the time period. The Paper Figures 8 show the same for Virality (retweets) and Popularity (likes) as a robustness check.

E.3 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 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 either 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 a 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 discretional but we argue it is a pretty conservative choice.

There is little overlap between the two definitions, hopefully indicating that we capture different kinds of bots successfully. Among our relevant tweets, a total of 16,700 tweets were identified through definition 1., a total of 5,748 through definition 2., while only 115 overlap. As a reminder we define relevant tweets 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 first concern is that bots may have inflated the presence of certain narratives in the dataset. Therefore, our first check examines the distribution of narratives. Table E.3 corresponds to paper Table 2, but accounts for potential bot activity by dropping tweets potentially posted by bots. It reports the share of character-roles and neutral instances for each character of interest, excluding those character-role that did not reach at least 100 occurrences in the full period, in line with the paper's table.

The results suggest that bots did not disproportionately promote specific narratives. Removing all (potentially) bot-generated tweets does not substantially change the composition of narratives. Fossil Industry–Villain and Green Tech–Hero remain the most common narratives. The share of Green Tech–Hero drops slightly once bot activity is excluded, which may suggest some degree of coordinated posting. However, all other shares remain virtually unchanged.

We take a step further to strengthen these descriptive findings by reproducing Figure 7, which presents the main results of the paper. Figure E.9 plots the coefficients from the same models used in the paper, but excludes tweets potentially posted by bots. The models apply increasingly restrictive specifications, moving from the baseline in the top panel to the most controlled model in the bottom panel. The results remain virtually unchanged, suggesting that bots do not play a central role in shaping the discussion on climate change policy, at least within the scope of our dataset.

Table E.3: ***Share of Character-Roles in Relevant Tweets, Excluding Potential Bots (United States, 2010-2021)***

Panel A: Human Characters

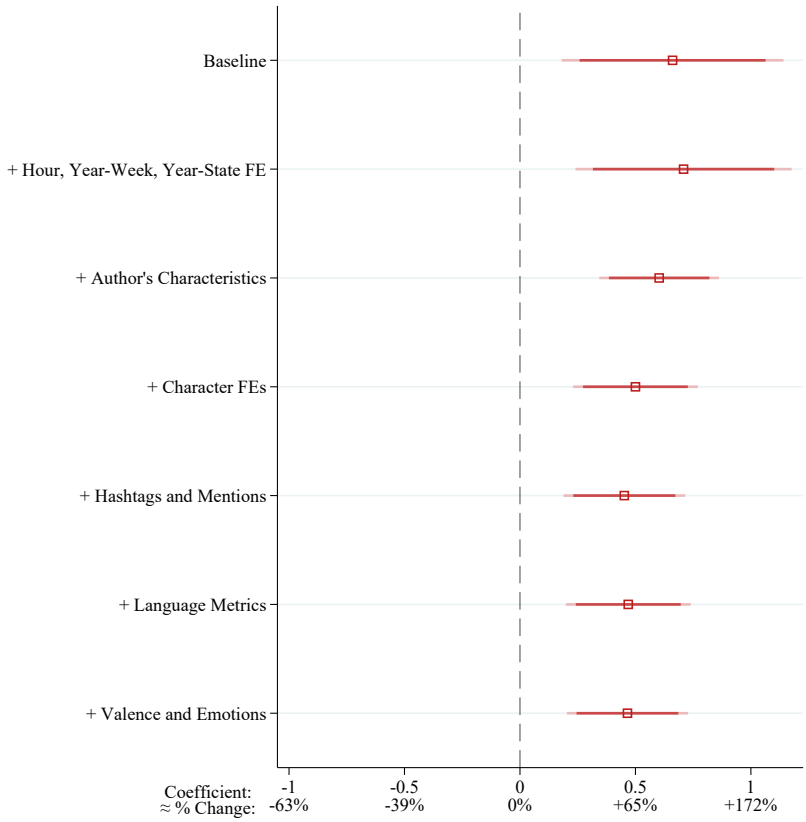
	Hero	Villain	Victim	Neutral	Total
Developing Economies	0.13	1.25	0.88	0.20	2.46
US Democrats	5.83	1.90	0.06	1.42	9.21
US Republicans	0.12	9.91	.	1.66	11.69
Corporations	0.96	8.10	0.06	7.78	16.90
US People	4.04	0.58	3.25	9.68	17.55

Panel B: Instrument Characters

	Hero	Villain	Victim	Neutral	Total
Emission Pricing	2.11	1.30	.	1.24	4.65
Regulations	3.51	1.44	.	3.14	8.09
Fossil Industry	0.07	11.48	0.04	1.67	13.26
Green Tech	10.41	0.88	.	3.16	14.45
Nuclear Tech	0.92	0.33	0.02	0.47	1.75

Notes: The table displays the frequency of character-roles in the classified data as shares of the total occurrences of each character. Shares are computed considering only 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 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 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. For this analysis we exclude tweets produced by potential bots, following the definitions provided.

Figure E.9: ***Regression Results - Impact of Political Narratives on Virality, Excluding Potential Bots***



Notes: The figure shows the coefficients of Poisson Pseudo-Maximum Likelihood regressions testing the effect of featuring Political Narratives as compared to featuring characters only in a neutral role, on measures of virality. The regression models include relevant tweets, defined as those featuring at least one character from our list. For this analysis we exclude tweets produced by potential bots, following the definitions provided above, and 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 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 baseline 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 model controls for tweet characteristics (number of hashtags and mentions). The sixth and seventh models include, respectively, language metrics and valence/emotions, as used in the descriptive analysis above. The figure reproduces the results of Paper Figure 7.

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 unimportant in shaping social media discussions more broadly. First, in recent years, significant improvements in bot programming may have enhanced their performance in ways not captured during our study period. Second, while climate change policy does not appear to be heavily affected, other topics — such as war, abortion, or general politics — may be much more vulnerable to bot activity. Finally, as noted above, there is no exact method for identifying bots, and the approach we use may have limitations in accurately detecting

automated accounts.

E.4 Alternative Regression Models

In this section, we want to dive deeper into the choice of model specifications used in our analysis. We explain the motivations and reasoning behind our decisions, while also considering potential alternative solutions. Our modeling choices are a response to the particular nature of the data and the research questions we address. In particular, the outcome variables of our analysis — retweets and likes — follow a power-law distribution: many observations take low values (including zeros), while a few take extremely high values. Such distributions are common in social media engagement and in other domains shaped by self-reinforcing processes. They pose unique challenges and require careful thinking when selecting an appropriate modeling strategy.

In similar cases, researchers often apply a log transformation to ease interpretation and reduce skewness before estimating models using Ordinary Least Squares (OLS). This would be a sensible solution if the dependent variables were strictly positive. However, our measures of Virality and Popularity — retweets and likes — contain a large proportion of zeros, which makes log transformations problematic.

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

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

In light of these issues, for this study, we adopt Poisson Pseudo-Maximum Likelihood (PPML) regression models as our preferred choice. Poisson models are particularly suitable for count data, especially when the underlying distribution follows a power-law structure, as is common in social media engagement metrics like retweets and likes. These models naturally handle zeros, avoiding the need for arbitrary transformations that could make results misleading. They also yield coefficients that can be interpreted similarly to semi-elasticities, maintaining a clear and meaningful link between model estimates and real-world effects.

In the remainder of this section, we present some alternative approaches. Although we argue above that Poisson represents the best method for our analysis, we provide these alternative models

to offer comparison and to improve understanding of the main results. In particular, we examine three approaches: OLS without any transformation, OLS applied after log-transforming the dependent variables, and the piecewise model proposed by Chen and Roth (2024), which distinguishes between extensive and intensive margin effects.

OLS Without Transforming

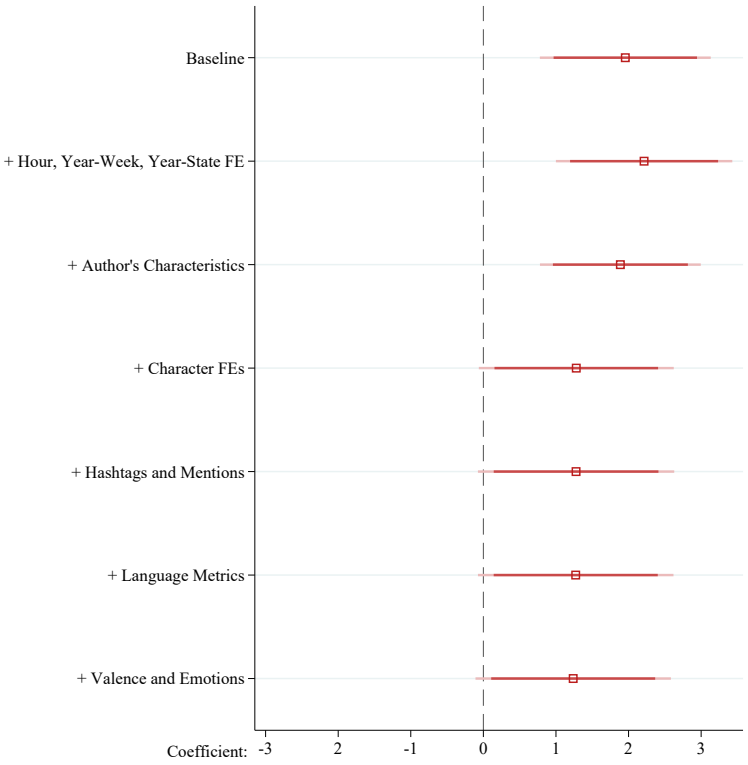
In the first part of this exercise we reproduce the results using simple OLS models. We do so leaving the dependent variable unvaried. The particular distribution of retweets and likes, following a power law, is problematic when using OLS and leads to a violation of the OLS assumptions, with regard to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

Figure E.10 reproduces the results of the paper Figure 7, where models are increasingly restrictive starting from the baseline model in the top panel, to the most restrictive one in the bottom panel. Overall the point estimates are in line with the main results of the paper. Featuring a Political Narrative increases the virality of tweets relative to tweets only featuring characters in a neutral framing. To add on what is said above, even if the point estimate can still be unbiased, the standard errors likely are not. That is why we should take with a grain of salt the significance levels of these coefficients.

In the first part of this exercise, we reproduce the results using simple OLS models, keeping the dependent variable unchanged. The particular distribution of retweets and likes — following a power law — poses challenges for OLS. It violates key assumptions, especially regarding the distribution and variance of errors. Nevertheless, we provide this comparison to highlight the differences between OLS estimates and those from what we argue is the most appropriate modeling choice.

Figure E.10 replicates the results from Figure 7 in the paper, using increasingly restrictive models from the baseline (top panel) to the most controlled specification (bottom panel). Overall, the point estimates are consistent with the main results: featuring a Political Narrative increases the virality of tweets relative to tweets featuring characters only in neutral framing. While the point estimates from OLS may remain unbiased , the standard errors likely do not. Because OLS gives disproportionate weight to extreme values and cannot account for heteroskedasticity in power-law data, the significance levels of these coefficients should be interpreted with caution.

Figure E.10: ***Impact of Political Narratives on Virality - OLS Without Transformation***



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, and likes, Popularity (robustness). 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 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 baseline 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 model controls for tweet characteristics (number of hashtags and mentions). The sixth and seventh models include, respectively, language metrics and valence/emotions, as used in the descriptive analysis above.

OLS with Log Transformation

In the second part of this exercise, we adopt an approach that is very common in economic research: applying a log transformation. Because our dependent variables contain many zeros, we first add a positive constant before taking the logarithm. While this practice is common in economics, as discussed above, it can lead to problematic results. Adding a constant to the dependent variable effectively rescales the coefficient by an arbitrary value. Changing the value of this constant alters the point estimates of the models, leading to potentially unreliable results. Figure E.12 serves two purposes. First, both sub-figures aim to reproduce the results of Figure 7 from the paper. Second,

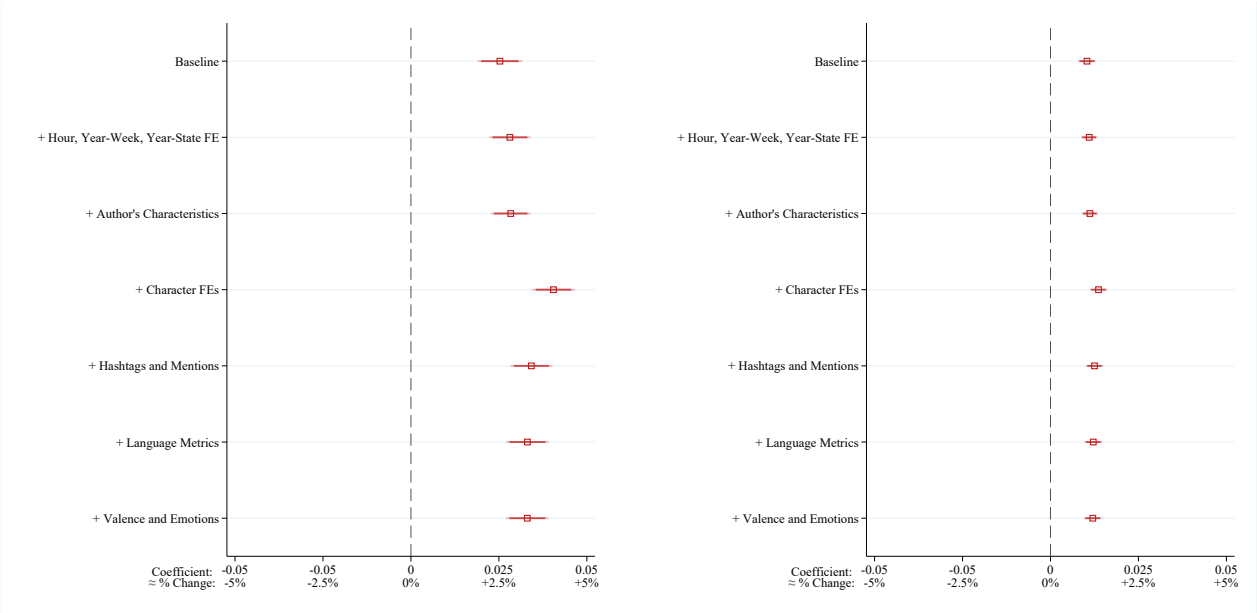
they illustrate the effects of adding different constants to the dependent variable before applying the log transformation to handle zeros. This allows us to demonstrate how changing the constant affects the estimated effects.

The first sub-figure on the left, Figure E.11a, applies a log transformation to $Y + 1$. As in previous models, specifications become increasingly restrictive from the top to the bottom panel, with each panel's header indicating the additional controls included. Several key differences emerge. First, for the Popularity measure, only the inclusion of user fixed effects makes the coefficients statistically significant. This highlights the importance of including user characteristics, as we do in the main specifications of the paper. Second, the percentage change implied by the logarithmic transformation is much smaller than the effects estimated using Poisson models. In the log-transformed models, the largest estimated increase is around 5%, compared to increases up to 160% in the Poisson models. The latter seems more consistent with the nature of the data. Nevertheless, the overall direction of the results remains in line with the main findings of the paper, providing an additional robustness check.

Figure E.11b also reproduces the results from Figure 7 in the paper, using log-transformed dependent variables. However, unlike Figure E.11a, this model adds 10 units to the dependent variable instead of 1 before applying the log transformation. If the transformation produced unbiased estimates, changing the constant should not affect the point estimates. Yet, in almost all models, the estimated effects are roughly half the size of those in the previous figure. Although the overall results still align with the main findings of the paper, this exercise highlights the weaknesses of log-transforming the dependent variable in this context. The instability of the estimates reinforces our argument that Poisson models provide a more reliable approach.

Figure E.11: ***Impact of Political Narratives on Virality - OLS and Log Transformation***

(a) Transformation: $log(Y + 1)$ **(b)** Transformation: $log(Y + 10)$



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, and Popularity (likes) as a robustness check. The dependent variable is log transformed after adding one unit, in Figure E.11a and 10 units, in Figure E.11b. 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 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 baseline 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 model controls for tweet characteristics (number of hashtags and mentions). The sixth and seventh models include, respectively, language metrics and valence/emotions, as used in the descriptive analysis above.

Extensive and Intensive Margin Model

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

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

For the second model:

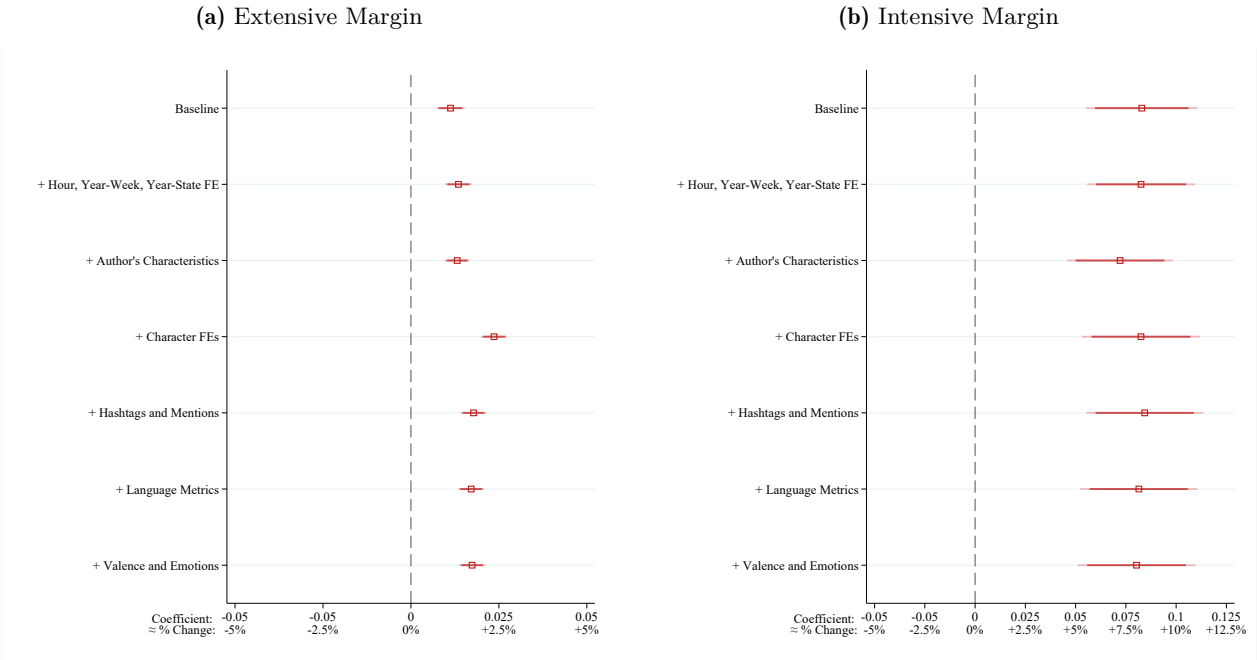
$$y^{*2} = \begin{cases} log(y) & \text{if } y > 0 \\ . & \text{if } y = 0 \end{cases}$$

The first model captures the extensive margin. It includes all observations but transforms any positive value of the dependent variable into 1, while leaving zeros unchanged. This approach captures whether featuring narratives increases the likelihood of any engagement with a tweet. By using OLS, we effectively estimate a Linear Probability Model, interpreting the results as the extensive margin of virality.

The second model captures the intensive margin. It retains only observations where the dependent variable is greater than zero. For this subset, we apply a log transformation to the dependent variable and estimate the model using OLS. This approach addresses a different question: Does featuring narratives increase or decrease the intensity of virality, conditional on receiving some engagement? In this model, the coefficients can be interpreted as semi-elasticities.

Figure E.12a and Figure E.12b show the extensive and intensive margin effects, respectively. The results point to a clear pattern. In both models, narratives have a positive impact on virality. Featuring a narrative increases both the likelihood of any user engagement and the degree of engagement, which we consistently refer to as virality throughout this study. The only exception appears with Popularity in the extensive margin model, where the effect is negative unless author characteristics are included. Once again, author fixed effects play an important role in reducing noise and producing more reliable estimates. We therefore consider the inclusion of author controls essential.

Figure E.12: ***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, and likes as a robustness check. In Figure E.12a the dependent variable is first transformed to have $Y = 1$ if $Y > 0$, then a Linear Probability Model is applied. In Figure E.12b 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 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 baseline 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 model controls for tweet characteristics (number of hashtags and mentions). The sixth and seventh models include, respectively, language metrics and valence/emotions, as used in the descriptive analysis above.

Summary

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

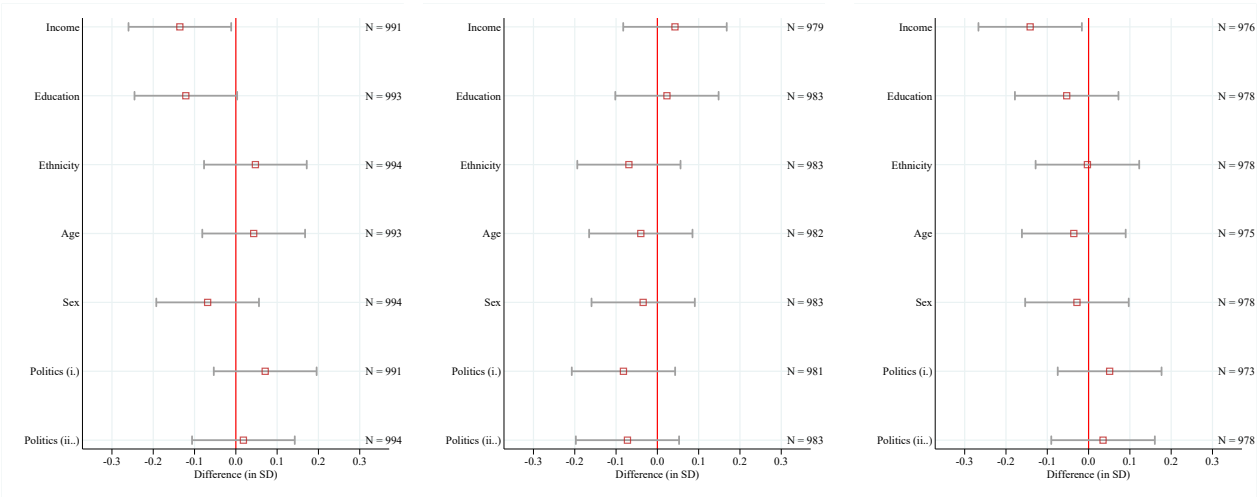
F Additional Output: Experimental Data

F.1 Additional Descriptive Output

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

F.2 Balance Check

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



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

Notes: The Figure shows control-treatment balance tests for several individual characteristics, among the participants of each experiment. Panel (a) shows result 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 characteristics on a dummy variable that takes value 1 if the participant is in the treatment group. We use robust standard errors.

G Robustness Checks: Experimental Data

G.1 Output Excluding Controls

Table G.1: ***Manipulation Check - Effectiveness of the Political Narrative Treatment - Excluding Controls***

Dependent Variable	**Mention of Character-Role**							
	Coeff./SE/p-value							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Political Narrative vs. No Role	0.296 (0.030) [0.000]	0.199 (0.032) [0.000]	0.118 (0.037) [0.001]	0.162 (0.019) [0.000]	0.359 (0.033) [0.000]	0.244 (0.040) [0.000]	0.122 (0.044) [0.006]	0.192 (0.023) [0.000]
Outcome | Character					✓	✓	✓	✓
Experiment: Hero-Hero	✓			✓	✓			✓
Experiment: Hero-Hero-Villain		✓		✓		✓		✓
Experiment: Villain-Villain-Hero			✓	✓			✓	✓
Experiment FE				✓				✓
Mean Outcome Control Group	0.33	0.34	0.44	0.37	0.45	0.52	0.64	0.54
Observations	994	798	699	2286	744	569	514	1684

Notes: The table displays the coefficients of OLS regression models providing a manipulation check of the effectiveness of the narrative treatment. The check is done by exploring whether participants are more likely to mention the character-roles, defining feature of a narrative, when exposed to the narrative treatment rather than to the control. Columns 1, 2, 3, and 4 display the effect of being exposed to the narrative on the likelihood of mentioning the character-role. Columns 5, 6, 7, and 8 display the same conditional on mentioning the character. The outcome variable is always a dichotomous indicator of whether the participant mentions the character roles when asked to recall anything about the text they saw. Models do not include personal characteristics as control. We use robust standard errors. Paper Table 4 shows the same table including personal characteristics as controls.

Table G.2: ***Experimental Results - The Impact of Political Narratives on Beliefs - Excluding Controls***

Dependent Variable	**Expectation for the Future:**					
	Forecast	**Confidence**	**Forecast**	**Confidence**	**Forecast**	**Confidence**
	Coeff./SE/p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Political Narrative	1.838	3.018	0.755	0.037	-4.995	-0.856
vs. No Role	(1.321)	(1.694)	(1.360)	(1.716)	(1.311)	(1.717)
	[0.165]	[0.075]	[0.579]	[0.983]	[0.000]	[0.618]
Experiment: Hero-Hero	✓	✓				
Experiment: Hero-Hero-Villain			✓	✓		
Experiment: Villain-Villain-Hero					✓	✓
Mean Outcome Control Group	39.02	46.02	43.67	48.54	42.32	48.15
Observations	994	994	983	983	978	978

Notes: The table displays the coefficients of OLS regression models providing insights on two outcomes: the participants' forecast, as answer to the question *'What percentage of U.S. 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 U.S. energy will come from renewable sources and green technology. How certain are you that the actual share of renewable energy in 2035 will be between [x-5] and [x+5]%?'* (in columns 2, 4, 6). Columns 1 and 2 show results for the Hero-hero experiment, columns 3 and 4 for the Hero-hero-villain experiment, and 5 and 6 for the Villain-villain-hero experiment. The models do not include personal characteristics as controls. We use robust standard errors. The Paper Figure 10a is the reference coefficient plot, where regressions also include controls.

Table G.3: ***Experimental Results - The Impact of Political Narratives on Stated Preferences - Excluding Controls***

Dependent Variable	**Stated Preferences: Policy Support**		
	Coeff./SE/p-value		
	(1)	(2)	(3)
Political Narrative	0.040	0.052	-0.011
vs. No Role	(0.092)	(0.092)	(0.126)
	[0.664]	[0.568]	[0.930]
Experiment: Hero-Hero	✓		
Experiment: Hero-Hero-Villain		✓	
Experiment: Villain-Villain-Hero			✓
Mean Outcome Control Group	5.70	5.72	4.19
Observations	994	983	978

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

Table G.4: ***Experimental Results - Impact of Political Narratives on Revealed Preferences - Excluding Controls***

Dependent Variable	**Revealed Preference: Incentivized Donation**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Political Narrative	0.540	0.570	0.311	0.474
vs. No Role	(0.477)	(0.458)	(0.469)	(0.270)
	[0.257]	[0.214]	[0.507]	[0.079]
Experiment: Hero-Hero	✓			✓
Experiment: Hero-Hero-Villain		✓		✓
Experiment: Villain-Villain-Hero			✓	✓
Experiment FE				✓
Mean Outcome Control Group	6.52	6.40	6.73	6.55
Observations	994	983	978	2955

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

Table G.5: ***Experimental Results - The Impact of Political Narratives on Memory (Pooled Sample) - Excluding Controls***

Dependent Variable	**Information Retention:**			
	Facts		**Character**	
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Political Narrative	0.001	-0.002	0.059	0.053
vs. No Role	(0.012)	(0.012)	(0.016)	(0.021)
	[0.917]	[0.847]	[0.000]	[0.011]
Experiment FEs	✓	✓	✓	✓
Day of the Experiment	✓		✓	
Day After the Experiment		✓		✓
Mean Outcome Control Group	0.13	0.13	0.69	0.45
Observations	2955	2297	2955	2286

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 regressions include experimental FEs and the following controls: income, income, education, political preference, age, and sex. SWe use robust standard errors. The Paper Table G.14 shows the same specifications, but including controls.

G.2 Output with Randomization Inference

Table G.6: ***Manipulation Check - Effectiveness of the Political Narrative Treatment - Randomized Inference***

Dependent Variable	**Mention of Character-Role**							
	Coeff./SE/ *Randomized (n=1000)* p-value							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Political Narrative	0.293	0.281	0.254	0.271	0.358	0.322	0.280	0.316
vs. No Role	(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]
Outcome | Character					✓	✓	✓	✓
Experiment: Hero-Hero	✓			✓	✓			✓
Experiment: Hero-Hero-Villain		✓		✓		✓		✓
Experiment: Villain-Villain-Hero			✓	✓			✓	✓
Experiment FE				✓				✓
Mean Outcome Control Group	0.33	0.34	0.44	0.37	0.45	0.52	0.64	0.54
Observations		976	968	2931	742	686	695	2123

Notes: The table displays the coefficients of OLS regression models providing a manipulation check of the effectiveness of the narrative treatment. The check is done by exploring whether participants are more likely to mention the character-roles, defining feature of a narrative, when exposed to the narrative treatment rather than to the control. Columns 1, 2, 3, and 4 display the effect of being exposed to the narrative on the likelihood of mentioning the character-role. Columns 5, 6, 7, and 8 display the same conditional on mentioning the character. The outcome variable is always a dichotomous indicator of whether the participant mentions the character roles when asked to recall anything about the text they saw. Models do not include personal characteristics as control. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. Paper Table 4 shows the same table including personal characteristics as controls.

Table G.7: ***Experimental Results - The Impact of Political Narratives on Beliefs - Randomization Inference***

Dependent Variable	**Expectation for the Future:**					
	Forecast	**Confidence**	**Forecast**	**Confidence**	**Forecast**	**Confidence**
	Coeff./SE/*Randomized (n=1000)* p-value					
	(1)	(2)	(3)	(4)	(5)	(6)
Political Narrative	2.487	3.426	0.831	0.331	-5.022	-0.358
vs. No Role	(1.368)	(1.680)	(1.425)	(1.740)	(1.387)	(1.803)
	[0.055]	[0.040]	[0.587]	[0.864]	[0.001]	[0.857]
Controls	✓	✓	✓	✓	✓	✓
Experiment: Hero-Hero	✓	✓				
Experiment: Hero-Hero-Villain			✓	✓		
Experiment: Villain-Villain-Hero					✓	✓
Mean Outcome Control Group	39.02	46.02	43.67	48.54	42.32	48.15
Observations	987	987	976	976	968	968

Notes: The table displays the coefficients of OLS regression models providing insights on two outcomes: the participants' forecast, as answer to the question *'What percentage of U.S. 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 U.S. energy will come from renewable sources and green technology. How certain are you that the actual share of renewable energy in 2035 will be between [x-5] and [x+5]%?'* (in columns 2, 4, 6). Columns 1 and 2 show results for the Hero-hero experiment, columns 3 and 4 for the Hero-hero-villain experiment, and 5 and 6 for the Villain-villain-hero experiment. The models do not include personal characteristics as controls. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. The Paper Figure 10a is the reference coefficient plot.

Table G.8: ***Experimental Results - The Impact of Political Narratives on Stated Preferences - Randomized Inference***

Dependent Variable	**Stated Preferences: Policy Support**		
	Coeff./SE/*Randomized (n=1000)* p-value		
	(1)	(2)	(3)
Political Narrative vs. No Role	0.061 (0.090) [0.495]	0.054 (0.086) [0.539]	0.094 (0.117) [0.410]
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 A, 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 B 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. Column 1 shows results for the Hero-hero experiment, column 3 for the Hero-hero-villain experiment, and column 5 for the Villain-villain-hero experiment. All models include income, education, political preference, age, and sex as controls. The models do not include personal characteristics as controls. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification. The Paper Figure 10b is the correspondent coefficient plot.

Table G.9: ***Experimental Results - Impact of Political Narratives on Revealed Preferences - Randomized Inference***

Dependent Variable	**Revealed Preference: Incentivized Donation**			
	Coeff./SE/*Randomized (n=1000)* p-value			
	(1)	(2)	(3)	(4)
Political Narrative vs. No Role	0.734 (0.497) [0.126]	0.530 (0.469) [0.252]	0.584 (0.480) [0.213]	0.562 (0.272) [0.046]
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	2931

Notes: The table displays the results from OLS regression models analyzing the impact of the narratives on participants' revealed preferences. We measure revealed preferences with the decision to donate to an association promoting sustainable development and local/national projects to support Green tech diffusion. The decision is incentivized with a lottery: participants could be selected to win 25$ and had to allocate the amount between themselves and the association. No restrictions on the allocation were given. Columns 1, 2, and 3 show results for the single experiments, column 4 shows results for the pooling together all experiments, and it includes experiment fixed effects. The models do not include personal characteristics as controls. We compute the p-value using the STATA command *ritest* for randomized inference, using 1000 repetitions, with *strict* and *two-sided* specification.

Table G.10: ***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)
Political Narrative	0.001	-0.002	0.061	0.049
vs. No Role	(0.012)	(0.013)	(0.016)	(0.021)
	[0.953]	[0.918]	[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	2931	2280	2931	2269

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 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. The Paper Table G.14 shows the same specifications, but including controls.

G.3 Additional Details on Experimental Output

Table G.11: ***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)
Political Narrative	2.487	3.426	0.831	0.331	-5.022	-0.358
vs. No Role	(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: the participants' forecast, as answer to the question *'What percentage of U.S. 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 U.S. 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. The Paper Figure 10a is the correspondent coefficient plot.

Table G.13: ***Experimental Results - Impact of Political Narratives on Revealed Preferences***

Dependent Variable	**Revealed Preference: Incentivized Donation**			
	Coeff./SE/p-value			
	(1)	(2)	(3)	(4)
Political Narrative	0.734	0.530	0.584	0.562
vs. No Role	(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	2931

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 the following controls: income, education and politics. SEs are clustered at the participant level. The Appendix Table G.4 shows the same specifications, but excluding controls.

Table G.12: ***Experimental Results - The Impact of Political Narratives on Stated Preferences***

Dependent Variable	**Stated Preferences: Policy Support**		
	Coeff./SE/p-value		
	(1)	(2)	(3)
Political Narrative	0.061	0.054	0.094
vs. No Role	(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. In column A, 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 B 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. Column 1 shows results for the Hero-hero experiment, column 3 for the Hero-hero-villain experiment, and column 5 for the Villain-villain-hero experiment. All models include income, education, political preference, age, and sex as controls. We use robust standard errors. The Paper Figure 10b is the correspondent coefficient plot.

Table G.14: ***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)
Political Narrative	0.001	-0.002	0.061	0.049
vs. No Role	(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	2931	2280	2931	2269

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 regressions include experimental wave FE and the following controls: income, education and politics. SEs are clustered at the participant level.

H Political Narratives in Other Media

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

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

H.1 Newspapers

In preparing and classifying the newspaper articles we follow as close as possible the procedure to prepare the tweets. We source a random and representative set of newspaper articles. To that end, we use the three most widely circulate newspapers in the US; The New York Times, The Wall Street Journal, and USA Today. We download the articles from Factiva. For each newspaper, we download 3000 articles for the period between 2010 and 2021. We download the articles in 4 year intervals. To ensure a balanced distribution of popular articles over time for each newspaper, we source the 333 and 334 most popular articles from 2010-2013, then 2014-2017, and 2018-2021. We use exactly the same list of keywords (Oehl, Schaffer, and Bernauer 2017)(see Section A.1 for the full list of keywords). We minimally adapt the prompts to classify the tweets, changing the references from tweets to newspaper articles. The adapted prompts can be found attached.

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

This resembles the ranking for instrument characters.

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

Panel A: Human Characters

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

Panel B: Instrument Characters

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

Notes: The table shows the frequencies of character-roles in the classified newspaper articles as a percentage of the total occurrences of each character. Shares are computed considering only the dataset of relevant snippets used in our analysis. We define a tweet 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 tweet but is not depicted in one of the three specific roles. The occurrence of character-roles is not mutually exclusive, meaning multiple roles may appear in the same tweet. The appendix Table C.6 shows the frequency in absolute numbers, including also the categories excluded here, because not reaching the 100 instances in the time period.

H.2 Television

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

Panel A: Human Characters

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

Panel B: Instrument Characters

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

Notes: The table shows the frequencies of character-roles in the classified tv transcripts as a percentage of the total occurrences of each character. Shares are computed considering only the dataset of relevant tv transcripts used in our analysis. We define a snippet as relevant if it features at least one character from our list. We include in the computation of shares only character roles that appear at least 100 times, thus excluding 'US Republicans-victim', 'Emission Pricing-victim', 'Regulations-victim', and 'Green Tech-victim', indicated by a dot in the tables. Panel (a) displays the shares for characters of the human type, while Panel (b) displays the same for characters of the instrument type. The column Neutral in both panels reports cases where the character is present in the tweet but is not depicted in one of the three specific roles. The occurrence of character-roles is not mutually exclusive, meaning multiple roles may appear in the same tweet. The appendix Table C.6 shows the frequency in absolute numbers, including also the categories excluded here, because not reaching the 100 instances in the time period.