Competing Imaginaries and Partisan Divides in the Data Rhetoric of Advocacy Organizations

ANONYMOUS AUTHOR(S)*

Data are wielded to shape public opinion, particularly in electoral contexts where the role and veracity of information is questioned. This post-truth era is characterized by world events in which facts too often are obfuscated and evidential standards are abandoned. To study how data are used to influence pressing and divisive contemporary issues, this paper explores the rhetorical work that quantitative data are doing through the blogging practices of advocacy organizations during the highly-polarized month preceding the 2016 United States elections. We present results of a qualitative content analysis of the quantitative data used in 337 blog posts published by five pairs of conservative and liberal advocacy organizations over the course of the month leading up to the 2016 US elections. We identify key data rhetoric practices along partisan lines and contribute an analytic framework–evaluating ethos, pathos, and logos— that can be used to analyze the rhetorical use of data in other contexts. We then characterize two different imaginaries that come into conflict in this research: 1) the political imaginaries being promoted through organizational blogging and 2) the sociotechnical imaginary of the data economy, foregrounding differences in the epistemic value of data in each. We conclude by outlining research challenges and trajectories for future research within each of the two imaginaries of data.

CCS Concepts: • **Advocacy**; • **Politics**; • **Misinformation**;

Additional Key Words and Phrases: Data Rhetoric, Advocacy Organizations, Post-Truth, Political Action Committee, Misinformation, Disinformation, Imaginaries, Political Blogs, Politics of Data

ACM Reference Format:
Anonymous Author(s). . Competing Imaginaries and Partisan Divides in the Data Rhetoric of Advocacy Organizations. In ,. ACM, New York, NY, USA, 28 pages.

1 INTRODUCTION

Data are not neutral [26]. Data, particularly quantitative data, possess both epistemic weight [27], as well as affective weight [95], and are used to legitimize arguments [24]. Understanding whether and how data are wielded to shape public opinion is crucial in a time when data too often are obfuscated, evidential standards are abandoned [84, p 1], and most critically, when the role of data, in general, is questioned [10].

Yet, understanding how data are used to shape public opinion requires more than an assessment of the extent to which data may be misleading and the intentionality behind it. It requires understanding the situated nature of data use [53], a phenomenon that Khovanskaya and Sengers have termed "data rhetoric" [61]. In the following excerpt, for example, the National Rifle Association isn't merely informing its audience about the number of people on the US "No Fly list," or about the gun purchasers—it is situating these data in a rhetorical context in order to discredit a single individual.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

> *"As of 2014, about **50,000 people were on the No Fly list**. This is a ten-fold increase since Barack Obama became president. Between February 2004 and December 2014, over **2,000 people** on the list purchased guns. Yet, not one of these people has been identified as using a gun in a crime." (National Rifle Association (NRA), 21 October)*

Other data are communicated more vaguely, through narrative language instead of specific numbers: "all families" or "every decision," for example,

> *Hillary Clinton is the **most qualified candidate we've ever had** running for president — and she's motivated by her compassion **for all families**, which influences **every decision she makes**. I am proud to be with Hillary because ever since I was a young woman in college watching her as first lady, she's shown me what it looks like to never give up or stop fighting for change — from her fight to get more than **eight million kids** covered under CHIP to her speech on women's rights to standing with New Yorkers as they re-built their city," EMILY's List President Stephanie Schriock (Emily's List (EMILY), 15 October)*

Data rhetoric are used to draw attention to different things of value, for example counting people on the "No Fly list" (NRA, October 21, 2016) or counting "families," "decisions," and "kids" (Emily's List, October 15, 2016). Data rhetoric are also used to create affect, whether negative affect as in the National Rifle Association example or positive affect, as in the Emily's List example.

Recent research has highlighted the "affective weaponization of information" in contemporary political propaganda [24, p. 2]. That is, political communication increasingly leverages data and its connotations of irrefutable ground truth as part of a stark play into the emotional manipulation of audiences. Multiple scholars [24, 56, 61] have called for qualitative analyses of how affective uses of language in political contexts manifest in digital environments:

> *It is becoming increasingly important to better understand the kind of environments facilitated by new media technologies, and their affective affordances: the rhythmic, habitual feed of signals and triggers, the cycles of outrage and laughter, the pleasure of "destroying" one's enemies in the name of Reason and civility.* [p. 98] [56]

Research about the rhetorical use of data represents a critical thread of scholarship in computer-supported cooperative work. Khovanskaya and Sengers [61] brought attention to data rhetoric through their case study tracing how the International Ladies' Garment Worker Union successfully conducted contentious data work. Erete et al. [44] took a more designerly approach to understanding data rhetoric; through their design of data infrastructures for nonprofits, they came to better understand the extent to which nonprofits preferred unprocessed data that could be flexibly reappropriated for the different storytelling needs of their partner organizations. Implicit here is the observation that different data rhetoric is likely to be charismatic to different audiences, drawing attention to certain phenomena over others. Multiple scholars have emphasized the extent to which quantitative data are a generally charismatic genre of information [26, 69], particularly in the nonprofit sector [13, 25]. Other researchers have conducted qualitative field work to more specifically characterize what data are more charismatic to what specific audiences [14, 95, 120].

Yet, audiences are often unaware of the various ways the information they see has been "cooked" [49] and "assume the information they are being presented with is representative of the broader universe of data that exists" [37]. Our research, then, focuses on understanding the strategies and tactics of persuasive data rhetoric—how data are used and how they might be "cooked." We undertake this research in the high stakes context of the data rhetoric of political advocacy organizations, such as the National Right to Life or Emily's List. These organizations' missions are focused on persuasive communication, often in one essential policy area. The advocacy work of these groups often forces candidates to address specific issues as a litmus test, focusing–or,

some would say, distorting—campaigns through the lens of that single issue [82], significantly influencing partisan politics in the process.

During the 2016 US election cycle, these advocacy organizations, including 527 Groups and Political Action Committees (PACs), were the largest and second largest sector contributors in the elections respectively, contributing over $153,760,471 [6]. Although the amount of money spent on elections is not an exclusive metric of political influence [39], advocacy organizations are among the loudest institutional voices in politics, and are known for their intense style of lobbying [82].

As individuals and institutions develop increasingly sophisticated uses of data in politics and persuasion, particularly as those data are propagated via algorithmic curation to specific audiences on social media, it is important to establish a more nuanced and sophisticated understanding of how data are leveraged in rhetorical contexts and to what end, enabling researchers and designers to implement better socio-technical solutions for mitigating harm. It is this call that we take up in this research.

In what follows, we take up the research questions: What rhetorical work is done by data in the public messaging of political advocacy organizations? How do conservative and liberal advocacy organizations use data differently? We first review research about the use of data in persuasive communication as well as research at the intersection of politics, social media, and data use. We then describe our *qualitative* content analysis of *quantitative* data use in the blogs of five pairs of conservative and liberal advocacy organizations during the politically active month leading up to the 2016 US election and present our analytic framework for analyzing data rhetoric—evaluating ethos, pathos, and logos–that others can reappropriate for further research about data rhetoric in other contexts.

We share the results of our analysis, characterizing the prevalence and use of quantitative data by these advocacy organizations, generally, as well as by conservative and liberal organizations, comparatively. Finally, we discuss how differences between the data use of conservative and liberal organizations may be related to differences in worldview and reflect on the broader implications of our findings. Despite their differing worldviews, all organizations in this corpus use data rhetoric to promote their *political imaginaries*, or the "collective structure that organizes the imagination and the symbolism of the political" [31]. Yet the epistemic value of data in this imaginary is strikingly different from the value of data in the sociotechnical imaginary of the data economy. As such, we conclude by exploring how these two imaginaries suggest different challenges and trajectories for future research.

2 REVIEW OF LITERATURE

2.1 Data in Persuasive Communication

In their book about statistics in the modern era, Hacking and Hacking note that facts are often presented in the context of propositions, which "can be assessed as true-or-false only when there is some style of reasoning and investigation that helps determine its truth-value" [53, p. 7]. Khovanskaya and Sengers similarly focus on the importance of understanding data use in situ [61]. They explored how how US labor unionists used data-driven rhetorical arguments to "bolster the legitimacy of organized labor's intervention." When the management and union brought in engineers to track the garment workers' micro-movements, the union was able to rhetorically frame these data in ways that highlighted the "mutual gains" of both management and workers, bolstering their own negotiating power while ultimately improving overall efficiency. These findings led Khovanskaya and Sengers to coin the phrase *data rhetoric* and to advocate for more research to explore the situated use of data, a call that we take up in this research [61, p. 1396].

In communication, framing is considered a key rhetorical mechanism by which people influence audiences. According to Entman, framing "is to select some aspects of a perceived reality and make them more salient in a communication text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation" [43, p. 52]. Numerous scholars have focused on "media frames," (e.g., how political messaging is introduced) and their influence on public opinion [52, 62, 102]. Studies highlight the importance of framing in political issues [59, 103], from analyzing preference toward the terms "climate change" or "global warming," to the language of vaccination rollouts. In science and technology studies, Bowker makes a similar case about quantification; that is, not only does language choice matter, but what is counted or measured matters [26]. Baumer et al. [22] posit that drawing attention to *political framing*—rhetoric that seeks to influence readers' opinions in overt and subversive ways on political matters—can help reduce its effects by enabling critical examination.

In political contexts, persuasive communicators control narratives by strategically using rhetorical frames to deliver facts [68]. Persuasive arguments frame problems in favor of one's political leaning [117], forcing any opposition to have to address both the issue in addition to its frame [18]. Studies of framing in political communication have found that conservative lawmakers use more negative language than their liberal counterparts, whether on websites [124] or on Twitter [115] or in print media [70]. Perhaps this is due, at least in part, to findings that the use of negative language in political messaging has been shown to garner more attention than positive messaging [71, 88]. While ample research has found evidence of partisan linguistic divides in language use, research has not yet explored whether these partisan divides extend to how conservatives and liberals use *data* in their communication.

2.2 Social Media and Politics

Politics has found a fertile ground for communication in online spaces since the 2004 US presidential elections when political blogs were widely used by both conservatives and liberals [11]. Adamic and Glance studied the prevalence of interlinking between political blogs in the 2004 elections, and found that linking behavior occurred primarily between ideologically similar blogs and that linking often occurred to discredit individuals. The 2008 US elections featured over 1,000 user-created Facebook pages pertaining to elections [126]. Since then, social media platforms continued to grow in popularity and algorithmic sophistication, allowing for media to be customized for niche interests and engagement [23]. In order to foster audience engagement, platforms such as Twitter privilege "discourse that is simple, impulsive, and uncivil" [92]. Advocacy organizations have joined their audiences online, using social media as new frontiers for their causes, funding [90], and as sources for data [33].

In recent years, the many entities involved in political campaigns have exacerbated the rift between people with conservative and liberal values [45], with waves of populism spreading across the world [17, 105], fueled in part by algorithmic polarization [106, 109]. With increasing nuance, philosophers underscore that there is not just one source of polarization within social media and that different methods of communication resonate with different audiences [89]. "Epistemic bubbles," for example, refer to a social structure where relevant voices have been excluded. Members of epistemic bubbles tend to lack exposure to diverse ideas and points of view. Members of "echo-chambers," instead, are part of insulated networks where beliefs are amplified or reinforced through repetition among them. Audiences of echo-chambers "systematically distrust all outside sources" [89].

2.3 Politics and Data

Data are also inherently political, shaped during each stage of design, collection, processing, and interpretation [78]. People use analytic and interpretive processes before presenting data, which Bowker [49] refers to as "cooking" data.

When scholars study the role of data in politics, they often focus on its "backstage" (see [50]) uses in political organizations, such as micro-targeted advertisements about citizens [42, 75, 97] or political outcome predictions [16, 80]. Malakoff suggests that while backstage uses of data are primary drivers of policy and decision-making, "economic concerns, religious views, and ideological perspectives on the role of government" are other considerations increasingly valued above evidence [77].

Studies of data use in "frontstage" politics, or the public-facing uses of data, have historically focused on data visualization, for example the diagrams used by the Ross Perot 1992 US presidential campaign [122] or the use of "narrative visualization," an approach to storytelling through data-driven graphical visualizations [104, 114]. Du et al. [40] found that the most effective presentations of data in the news were associated with data visualizations; yet they also argue that the utility of data in convincing audiences has subsided as audiences have been increasingly inundated with facts and statistics [40].

Boler and Davis also posit that data have become less charismatic in our post-truth era, though they argue that this is because data are rational and do not elicit strong emotions [24]. Hong, instead, interrogates the role of emotions in the use of data by tracing how "charismatic influencers" leverage facts in order to seem like they care about facts and reason [56]. This practice, which Hong terms calls *fact signaling*, imbues data with "the strategic and performative invocation of epistemic and moral authority which may then be weaponized" [56, p. 86]. Similarly, Bakir and McStay [19] argue that emotions are leveraged in order to attract an audience's time and attention, and that facts take on symbolic roles to add legitimacy.

Studies that investigate narrative forms of frontstage data often assess the acceptance and buy-in of mis/disinformation in this post-truth era (e.g., [29, 93]) and the role of *empathic media*, or media that captures the emotional attention of audiences, in spreading that disinformation (e.g. [19, 20, 118]). This body of research highlights how alternative media ecosystems are designed to "undermine trust in information generally" rather than spreading a specific ideology [96]. Misinformation is a byproduct of *disinformation*, which is a purposeful effort to spread wrongful or misleading information. When people unwittingly spread disinformation and believe it to be truthful, it is *misinformation* [34]. Calo et. al suggest that disinformation campaigns may not be easily discernible, and "may, ironically, involve true information and reasonable opinion" and recommend high level strategies to combat misinformation, such as relying on reputation and policy change [34, p 1]. Other themes in this body of research explore the susceptibility of populations to mis/disinformation [94], model how lies propagate [110, 112], and explore the intentionality behind content that contributes to post-truth practices [119]. In considering how to more effectively stop the spread of disinformation and to help people spot misinformation, researchers have suggested both design interventions [111] and policy interventions [34].

3 METHODS

3.1 Research Context

3.1.1 Advocacy Organizations and Longform Communication. Blog posts, like email listservs, are widely used in electoral communication to spread appeals to dedicated members of a specific community. In the context of US politics, most studies of blogs occurred between the 2004 and 2012 elections cycles (e.g. [11, 23]), focusing on understanding the breadth of distinct topics discussed

on such forums [11], To this day, political organizations still leverage blog posts and emails with their constituencies to conduct work including: poll tracking [87], fundraising (e.g., [86]), providing election coverage and spreading organization-centered news [38, 54]. Although they have smaller audiences, blogs remain a longstanding and necessary method of distributing information to constituents; their detachment from social media platforms and traditional media enable them to control their media environment [38].

Political blogs often serve as partisan spaces [72] and advocacy organizations' blog posts are often focused on the potentially-partisan missions of their organizations [90]. Many organizations also take advantage of microblogging platforms such as Twitter and Facebook, however, despite the larger audiences, microblogging is often a supplementary activity to an organization's publications; organizations will author longform content and then use their microblogging platforms to link back to that content, including blog posts [48]. Our analytic focus, then, is on the longform blog posts, as they offer more developed arguments and serve as the source for much of the data-driven content posted on microblogging platforms.

Studies of how other social media platforms were utilized during the 2016 elections show that organizations and campaigns leveraged such platforms most prominently for micro targeted advertising (e.g., [116]) and grassroots political organizing (e.g., [12]). Our survey of the Twitter and Facebook use of the highest spending single-issue advocacy organizations of the 2016 US election cycle reaffirmed that this specific subset of organizations also were still using these platforms for appeals including calls to action, reminders of events, requests to donate, participating in social media trends, and fulfilling other repertoires of action [35]. During the 2016 election cycle, political actors leveraged social media at levels of sophistication that were not seen prior to that election cycle [116]. However, there were few instances in which advocacy organizations used quantitative data in their Facebook or Twitter posts and when they did, they were most typically used in posts that linked directly to extended blog posts from which the data were excerpted. In the runup to the 2016 US elections, then, blogs were the social media platform on which the advocacy organizations in this study were most frequently and consistently using data rhetoric. As such, we focused our analysis on the use of data in organizations' blog posts.

3.1.2 Politics During the 2016 US Election Cycle. The 2016 US general election marked the ending of the Obama presidency and a stark change in communication strategies for political campaigns, with social media emerging as a more powerful mechanism for swaying public opinion [24]. The presidential race between Republican Party nominee, Donald Trump and Democratic Party nominee, Hillary Clinton became infamous for the immense amount of mis/disinformation that circulated over social media [21]. Donald Trump won the presidency in the general election by securing enough votes in the electoral college, although he lost the popular vote by a significant margin.

While Republican conservatives represented the challenger for the presidency, they had been the party in control of both the US House and Senate since 2014 (in the 114th United States Congress) with 246 seats out of 435 in the House and 54 seats out of 100 in the Senate. The Democratic liberals, then, were the challengers in these (aggregate) Congressional elections. In the end, the Republican conservatives retained their majority into the 115th United States Congress, though their lead narrowed to 241 seats and 52 seats respectively.

In state and local elections, there were twelve states with gubernatorial races, where Republicans were defenders in four states and Democrats were defenders in eight states [3]. Further, ten states held elections for attorney general [2], four states had ballot initiatives related to guns, and one ballot initiative related to life (assisted suicide) [5].

Much of US politics is dominated by attempts to sway single-issue voters—people who will cast a vote for a candidate based on one issue that is most important to them rather than a sum of the

candidate's entire platform [36]. Going into the 2016 election, there were a significant number of such single issues that both Republican and Democratic candidates fixated on [9, 100]. Such fixations on single issues can make them the objects of disinformation, such as the 2015 series of doctored videos targeting Planned Parenthood, which were circulated over social media in an attempt to discredit many stakeholders [73]. Political contenders seeking to win their elections stood to benefit greatly by aligning themselves with single-issue voter perspectives. In this research, we seek to better understand the persuasive strategies that single-issue advocacy groups employ in persuading people on casting their vote.

3.2 Constructing the Corpus: Selecting Advocacy Organizations

Advocacy organizations are considered particularly influential in mobilizing voters around single-issues such as women's reproductive rights [79], guns [66], immigration [98], and the environment [23]. Single-issue advocacy organizations are "known to run high-profile media campaigns to mobilize or recruit members" and make the most campaign contributions from any other class of interest group [8]. In the 2016 presidential election, Democratic candidate Hillary Clinton received $45 million from advocacy organizations, and Republican candidate Donald Trump received $4.9 million [7]. The goal of these organizations are to "speak for and mobilize broad constituencies" in an effort to influence a policy area [108]. We also included the public-facing communicative arm of each of the US political parties in order to gain insight regarding how more broad-based—but still influential—advocacy organizations use data.

We selected 10 nonprofit advocacy organizations Table 1 for this research based on a number of heuristics :

- We surveyed the social media ecosystems for both US political parties identifying the organizational arm responsible for public-facing blog posts—the Republican National Committee and the Democratic Political Party.
- The Center for Responsive Politics[1] tracks federal campaign contributions and spending. Referencing their schema of single-issue "industries," we identified the four constituency issues for which organizations represented both political ideologies: women's reproductive issues, guns, immigration, and the environment. The single-issues without ideological pairing (and so not included in this research) included: foreign & defense policy, pro-Israel, human rights, and LGBTQIA rights & issues. A full list of single-issue industries can be found in the appendix.
- For each single-issue topic, we identified the largest contributing organizations (aggregating expenditures of organizational affiliates, as appropriate) to the 2016 US election representing each ideology.

We then surveyed the social media ecosystems of each organization, confirming that these organizations consistently used more data rhetoric in their blogs than on other social media platforms. All organizations communicated to the public via a similar genre of written content, reflecting organizational commentary on current events; most labeled these as their blogs, a few titled their pages as "News" or "Updates" instead. In 2016, the blogs of advocacy organizations constituted an influential source of information and news, as their audiences include both mainstream media and public readerships [11, 46, 72, 85] and influence framing of the news [23, 48].

3.3 Constructing a Corpus of Organizational Posts

For each organization, we collected all text from blog posts (including transcripts of embedded videos) from one month prior to Election Day (8 October) through the day after the election (9

[1] https://www.opensecrets.org/

Table 1. Advocacy Organizations Represented in the Corpus

Issue & Organization	Ideology	Tax Code	$ in 2016 Election	# of Posts
Party				
Republican National Committee (RNC)	Conservative	527	$343,371,200	58
Democratic Party (DNC)	Liberal	527	$372,182,925	28
Reproductive Rights				
National Right to Life / NRL Victory Fund (NRL)	Conservative	527	$1,909,284	241[2](75)
EMILY's List/Women Vote! (EMILY)	Liberal	SuperPAC	$81,771,326	62
Guns				
National Rifle Association Institute for Legislative Action (NRA)	Conservative	501(c)(4), PAC	$54,398,558	24
Giffords PAC/Americans for Responsible Solutions (Giffords)	Liberal	Carey Committee	$13,468,557	15
Immigration				
Federation for American Immigration Reform (FAIR)	Conservative	501(c)(4)	$167,000	38
National Immigration Forum (NIF)	Liberal	501(c)(3), 501(c)(4)	$220,000	7
Environment				
Freedom Partners (funded by Koch Industries) (FP)	Conservative	PAC	$29,728,798	17[3]
NextGen Climate Action (NextGen)	Liberal	superPAC	$96,036,921	20

November), to include election result reactions. While organizations often had multiple websites, we analyzed the organizational website that served as the primary host of public-facing political communication, observing all blog posts published during the time-frame (with two exceptions noted in Table 1).

We did not analyze data presented in large blocks of text quoted from other sources (often news outlets), as these texts were not in the organizations' voice. Such posts were, however, included in both total post and word counts to demonstrate longitudinal posting behaviors. We did analyze

[2]The NRL posted 241 times, while no other organization posted more than 63 times. Together, the remaining conservative organizations posted 131 times combined. In order to prevent our conservative dataset from being characterized overwhelmingly by NRL posts, we calculated the average number of posts made by all conservative organizations (≈ 75) and divided the NRL sample size to select every third post in order to get an even distribution. This resulted in a sample of 80 NRL posts. We then dropped one post every 20 posts, leaving the final corpus of 75 posts.

[3]FP removed their website prior to data collection. We accessed their posts by using the Internet Archive's Wayback Machine. 17 of the 21 articles posted during the time period were available, which are included in our corpus.

blocks of reposted text from personal blogs; these were typically framed as being from guest contributors invited by the organization.

Our final corpus included 337 posts, 206 from conservative organizations and 131 from liberal organizations.

3.4 Data Analysis

While compiling the corpus, we constructed a spreadsheet of metadata, including a link to each post, the date posted, and the word count. We then conducted a series of content analyses across the corpus to identify themes and generate categories [63].[4]

3.4.1 Identifying Instances of Quantitative Data. In our content analyses, we first identified each instance of quantitative data which, for this rhetorical context, we operationalized as anything being counted. Most often, this meant the use of numbers, although sometimes the counting was implied or vague. For example, "DNC raises $36.6 million in September" and "2016 is on track to be the warmest year on record" would each be considered instances of quantitative data. We excluded numbers that provided temporal context instead of serving to count something. In the second example above, then, "2016" would not be considered quantitative data, because it serves as temporal context, but "warmest" would be considered a narrative description of quantitative data. We did not double count duplicate mentions of data made in the same sentence (e.g., "more than half (53%)").

3.4.2 Analytic Framework for Evaluating the Ethos, Pathos, and Logos of Each Instance of Quantitative Data. We employed multiple iterations of analysis to develop our coding scheme, which serves as an analytic framework for evaluating the ethos, pathos, and logos of each instance of quantitative data.

The first author read the first 10 posts in the corpus written by each organization, coding each instance of quantitative data using the guiding question, "What work is this instance of data doing?" The inductive open-coding during this initial phase of analysis led the authors to identify five analytic categories related to three Aristotelian rhetorical principles of persuasion: ethos, pathos, and logos. Ethos is the credibility of the argument. Here, for each instance of data: what is counted and with what degree of precision. Pathos is the connected emotion. Here, pathos focuses our analytic attention on whether the data is used with positive, negative, or neutral connotations—to credit or discredit an entity and what that entity is. Finally, logos is the determination of whether the logic underlying the argument is sound. For each instance of data, then, our analytic framework specifically asks:

Evaluating Ethos: *What Data Counts?* Determining what is being counted and how data are presented to readers reveals what entities may be deemed charismatic by the advocacy organization as well as what scope of information might be available to them.

- *Coding Category 1:* What is being counted (i.e., money, people, time, communicative reaches, votes, crime, legal death)?
- *Coding Category 2:* How are the data being communicated (i.e., specific number, calculation, narrative description, visually, or vague)?

Evaluating Pathos: *What are the Rhetorical Work of Data?* Identifying whether the instance of data presented is crediting or discrediting an entity reveals the rhetorical work that the data

[4]Design for this study began in 2018. Data collection commenced in 2019, with the first round of analysis lasting through summer 2019. Workshops and planning for further analysis occurred in fall of 2019 and the second round of analysis began in early 2020, lasting through summer 2020. Thanks to insightful feedback from anonymous reviewers, we expanded our corpus and conducted additional longitudinal analyses.

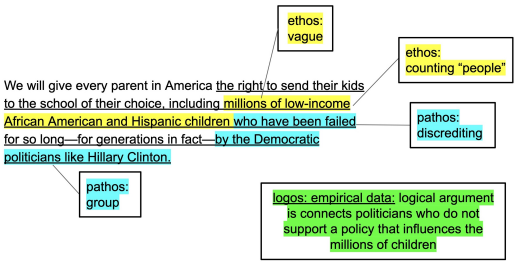

Fig. 1. RNC 05 November 2016.

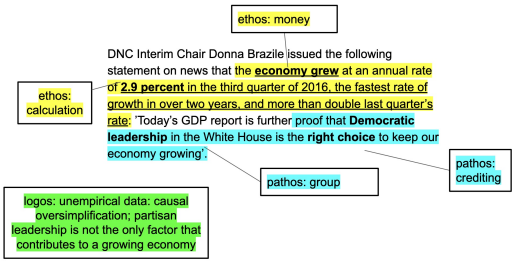

Fig. 2. DNC 28 October 2016.

do. Further, the entity that is being credited or discredited reveals the focus of the rhetorical work.

- *Coding Category 3:* Is this data making an issue or entity appear more positively, negatively, or neutrally (i.e., crediting, discrediting, or neutral)?
- *Coding Category 4:* What entity is the object of affective focus for the data (i.e., individual, organization, etc…)?

Evaluating Logos: *Are the Data Empirical?* By assessing the relationship between the instance of quantitative data and the surrounding text, we can better understand the rhetorical maneuvers taken by organizations to sway their audiences.

- *Coding Category 5:* Is this instance of data empirical (logically sound instance of counting) or not? If unempirical, in what way does it deviate from empirical data (i.e., projection of the future, numerical extrapolation, or unsound logical turn)?

Our inductive open-coding also resulted in sub-level codes for each of the five categories, which we foreshadowed above. To ensure a complete coding scheme, we engaged in a second round of analysis using MAXQDA[5], a text and multimedia analysis software for qualitative and mixed methods data. The first author read through the entire corpus, noting each instance of quantitative data and identifying the appropriate categories for the data according to the five categorical questions. The first and third authors, then, worked to refine the coding scheme, elaborating the set of sub-level codes for each categorical question and articulating initial definitions of each code.

In Figure 1 and Figure 2, we offer two examples of coded data. In Figure 2, the Democratic Party counts *money* as a percentage (or *calculation* to credit a specific group of people,, the Democratic leadership. The use of data is unempirical, however; the logic employs causal oversimplification; a single group of people cannot solely be responsible for an improved national economy.

In the case of Figure 1, the Republican National Committee counts *people* ("children") *vaguely* in order to discredit a *group* ("politicians") regarding a specific policy stance. Data are not merely used in order to convey information, but are additionally used in order to change readers' opinions about "Democratic politicians like Hillary Clinton." What is being counted (vague counts of people) and the entity these data are connected with (discrediting politicians with a specific policy stance that affects what is being counted) together present a sound logic argument.

Our analysis does not evaluate the truth-value of a given instance of data (i.e., we have not fact-checked each of the 2245 instances of quantitative data in our corpus); but rather, we evaluate the soundness of the logic in the argument in which data rhetoric is used. In doing so, we explore how facts, assumed to be true, are maneuvered to sway public opinion.

[5]MAXQDA.com

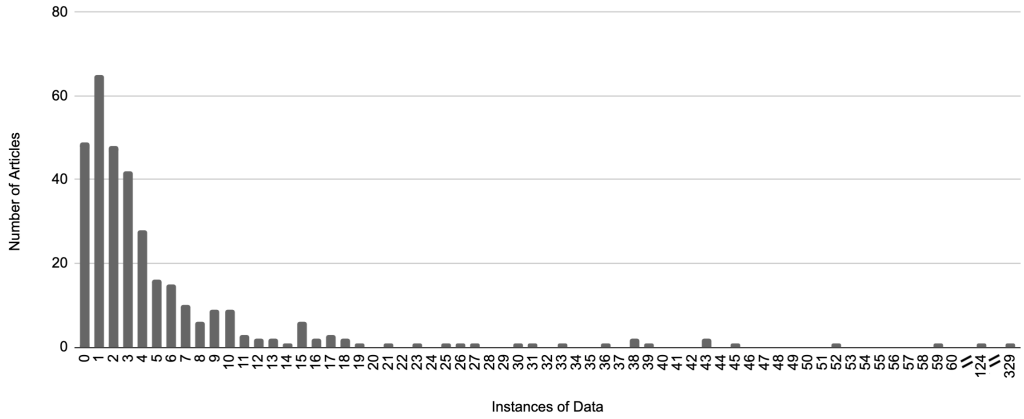

Fig. 3. Distribution of Data Across Posts

Following the first complete application of the coding scheme on the data, the authors further refined their codes in three workshops with 10 researchers who were all familiar with qualitative research. Workshop participants read selections from the corpus, including excerpts that were both ideologically diverse and representative of common instances of data encountered in the text. The selections also included two excerpts that were thorny instances of data use—edge cases that were used to prompt discussion and facilitate the development of additional nuance in the coding scheme. Feedback from these workshops resulted in adjustments for added precision to several definitions (e.g., data about murder counting as a 'crime', not a 'death' (re-named 'legal death')). The first author then completed another full pass through the data to apply this refined coding scheme.

The authors determined inter-rater reliability to reveal further definitional weaknesses and clarify the coding scheme [76, 83]. To calculate reliability, the first and second authors applied the coding scheme to a 10% sample of the corpus and compared results, calculating the Kuckartz and Rädiker Kappa (ignoring unassigned codes), with an agreement of $\kappa \geq 0.725$ established on all codes [64]. Disagreement was primarily due to instances of data not being defined a priori. Each coder differed slightly in the text segment length they defined as individual instances of data. The authors iterated on disagreements until clarification and agreement was reached, with new definitions subsequently applied to the full corpus by the first author.

4 RESULTS: HOW ADVOCACY ORGANIZATIONS USE DATA IN THEIR BLOGS

Across the 337-post and 234,107-word corpus, the 10 organizations use quantitative data 2,245 times. While the number of instances of data varies across posts, 288 (85%) of the posts in the corpus include at least one instance of data (Figure 3). On average, organizations use about seven instances of data per post, or one instance of data for every 104 words. Averages are skewed by a subset of posts in the long tail of the distribution with a much larger prevalence of data; nearly half of the posts contain one, two, or three instances of data (median = 3).

All organizations use some posts specifically for reporting data (e.g., candidate records, research results relevant to their issues, campaign efforts and successes, and organizational financial information) resulting in a subset of texts with high-intensity data use. Many of the data-dense posts had timing that co-occurred with news events including scandals as well as gubernatorial, congressional, or presidential debates.

Although the conservative and liberal organizations in our corpus include data at similar rates, the blog posts of conservative organizations are more frequent, lengthier, and use more data than their liberal counterparts' entries. The conservative and liberal organizations present an instance of data every 111 and 89 words, respectively. However, the conservative posts are significantly lengthier (168,202 words across 206 posts) than the liberal ones (65,905 words across 131 posts). As such, 67% of the total instances of data in our corpus are by conservative organizations (1,509 instances), and the other 33% are by the liberal organizations (736 instances). This disparity stemmed not only from the lengthier posts, but also from a distinct genre of data-intense posts written by the conservative organizations which aim to discredit politicians' records in office. The two most data-intense posts (329 instances on 14 October and 124 instances on 1 November), both from FP, both use data to review state-level lawmakers' records in office, quoting historical news coverage of the officials' votes and stances.

Data use by both conservative and liberal organizations in our corpus correlated with significant political events such as then-FBI director James Comey's announcement of investigation into Democratic presidential nominee Hillary Clinton's emails (28 October). Their data use also corresponded with key debates, most frequently Senatorial debates, though conservative and liberal organizations often posted about different debates. Liberal organizations, for example, posted about the Nevada (17 October) and New Hampshire (27 October and 2 November) senatorial debates while conservative organizations posted about the New York Congressional district debate (13 October), the third presidential debate (14 October), Wisconsin senate debates (14 October and 18 October), Indiana senate debate (18 October), Pennsylvania senate debates (17 October and 24 October), and Florida senate debates (17 October, 26 October). Conservative and liberal organizations also differed in the rhythm of their data use leading up to the election. The liberal organizations in our corpus used data more often as Election Day (8 November) approached, with small spikes in data use around the events noted above while conservative organizations used data more consistently throughout the month.

Due to the disparity in total instances of data use between conservative and liberal organizations, unless otherwise noted, we report either holistic numbers or the comparative rates of data use for all organizations of each ideology or when comparing data use between individual organizations. We do not, in general, report comparisons of data use between issue areas as our sample includes only one conservative and one liberal organization advocating for each issue.

4.1 Evaluating Ethos: What Data are Counted?

Understanding what is being counted provides evidence of what counts to these organizations. It helps gain insight to what is valued and what is believed to have "charisma" [26] for their audiences. Across the entire corpus, organizations counted money more than anything else (803 instances; 36% of total data). Yet, only two organizations, the conservative energy organization (FP) and the RNC, counted money (e.g., economic growth, donations, investments) more than anything else. However, they did so to an extreme that their counting of money masks a trend of eight of the ten organizations counting varying units of people (e.g., citizens, gun owners, students) the most (650 instances; 29% of total data). Other data counted time (e.g., terms in office or duration of pregnancy) (151 instances, accounting for 7% of total data), communicative reaches such as emails or phone calls, (115 instances; 5%), and votes (118 instances; 5%). 9% of the data (210 instances) count something that appeared even less commonly.

We also analyze the degree of specificity each instance of quantitative data are presented at, whether such data are presented as specific numbers (e.g., "average donation in September 2016: $63"), calculations (e.g., "homeownership fell to 62%"), vague (e.g., "the seven-figure paid media campaign"), or in descriptive forms (e.g., "the nation's fourth-highest enrollment rates"). This

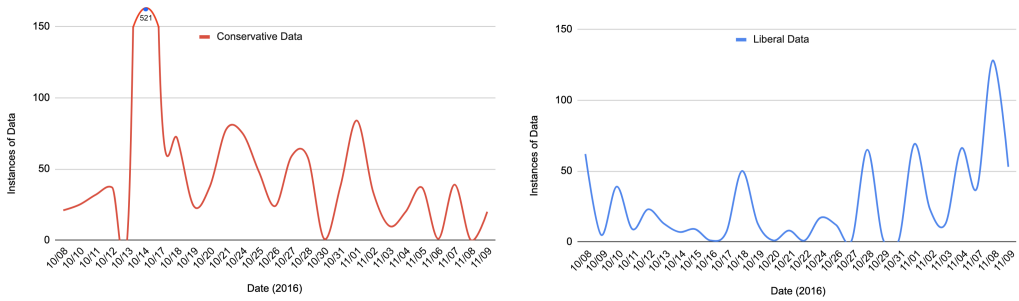

Fig. 4. Aggregate Use of Data by Conservative (Left) and Liberal (Right) Organizations

Table 2. Top five categories most frequently counted

	Conservative	Liberal	Entire Corpus
1	Money (678 instances; 45% of all conservative data)	People People (348 instances; 47% of all liberal data)	Money (803 instances; 36%)
2	People (302 instances; 20%)	Money (125 instances; 17%)	People (650 instances; 29%)
3	Time (90 instances; 6%)	Time (61 instances; 8%)	Time (151 instances; 7%)
4	Votes (84 instances; 6%)	Communicative Reaches (49 instances; 7%)	Votes (118 instances; 5%)
5	Communicative Reaches (66 instances; 4%)	Votes (34 instances; 5%)	Communicative Reaches (115 instances; 5%)

category was integral in determining what parts of the full corpus were coded. Our unit of analysis is instances of data. As some things may be too vague to be counted, we kept track of the spectrum of specificities and modalities of data to understand how the organizations evoked counts of data in their blog posts. Most frequently, data are presented as specific numbers (1085 instances; 48%), followed by calculated statistics (475 instances; 21%), vague data (437 instances; 19%), and narrative descriptions (272 instances; 12%). The rare instances in which visual data appear in the corpus (2 instances; 0.001%) are presented exclusively by conservative organizations. The specificity and modality of data may also suggest degrees of charisma and legitimacy, as well. Bowker, for example, attributes additional charisma to data as calculations and other formally presented information [26]. On the other hand, narrative descriptions may serve as opportunities to evoke more emotional responses.

4.1.1 What is Being Counted? All of the liberal organizations use data to count people (348 instances; 47% of liberal instances of data) more than anything else, and consistently do so at higher rates than their conservative counterparts. There was more variance in what conservative organizations counted. Together, the RNC and FP accounted for 640 of the total 678 conservative instances of counting money. These two organizations only accounted for 79 of the total 302 conservative instances of counting people. The other three conservative organizations focused on counting people, and did so 223 times; they only counted money 38 times. Notably, liberal organizations found ways to talk about financial issues that maintained their focus on counting people, for example counting donors instead of donations: “Millions of donors are supporting our candidates and our party because they believe we are stronger together” (21 October). This

Table 3. Distribution of different levels of data specificity

	Conservative	Liberal	Entire Corpus
Numbers	53% (799 instances)	39% (286 instances)	48% (1,085 instances)
Calculations	22% (335 instances)	19% (140 instances)	21% (475 instances)
Narrative Descriptions	7% (101 instances)	22% (162 instances)	12% (263 instances)
Vague	19% (284 instances)	21% (153 instances)	19% (437 instances)
Visual	0.001% (2 instances)	0 instances	0.0009% (2 instances)

deliberate use of human-forward language demonstrates how organizations across ideological rifts discuss the same topics, but use quantitative data as evidence in distinct linguistic ways.

The four organizations that counted people more than any other organizations were the four single issue advocacy organizations working in the areas of immigration (FAIR, NIF) and reproductive rights (NRL, EMILY). It may be, then, that some issue advocacy organizations find people-centric data more charismatic than organizations or audiences for other issues.

4.1.2 How Precisely are Data Being Counted? Conservative and liberal organizations in this corpus counted different things at different levels of specificity, as well. Both liberals and conservatives used specific numbers more than any other level of specificity, although conservatives relied on using specific numbers more heavily than the liberal organizations. 53% of all conservative data used specific numbers (799 instances), whereas 39% of all liberal data used specific numbers (286 instances). Across ideology, organizations used calculations and vague data at similar frequencies. More striking, the liberal organizations in this corpus used more narrative descriptions (22% of liberal data; 162 instances) than conservative organizations (7% of conservative data; 7%). Such evocations of counting served many purposes, notably, counting the milestones achieved by the election of endorsed candidates (e.g., "We've fought for women like…immigrant rights activist in Washington State, and Lisa Blunt Rochester, poised to be the first woman and person of color to serve in Congress from Delaware" (EMILY, 08 November)). Conservative organizations used visual data in two instances while liberal organizations did not use visual data at all.

Overall, specific numbers and calculations may be perceived by organizations as offering stronger validity than narrative descriptions or vague counts of data. The dearth of visual presentations of data are one of the most surprising findings of this research and the organizational rationale for this would be a compelling topic for future research.

4.2 Evaluating Pathos: What is the Rhetorical Work of Data?

The organizations we sampled from for this corpus used data as a rhetorical tool with affective implications. As such, we next take a broader view of the context surrounding each instance of data and its often political or persuasive work.

Organizations in this corpus rarely use data neutrally; the preponderance of data (2,144 instances; 96% of all data) *credit* or *discredit* an entity. The organizations most frequently use data for *discrediting*—making entities or issues appear more negatively (1,255 instances; 56%). They also use data for *crediting*; 889 instances of data (40%) are used to make entities appear more positively (Table 4). Only 5% (103 instances) of data across the entire corpus are neutral—neither crediting or discrediting (Table 4). The neutral instances of data consist primarily of administrative reporting (e.g., total cash on hand) or contextual information about policy and issue stances (e.g. how many lawmakers support a bill or eligibility details for policy benefits). The abundance of quantitative data with affective connotations emphasizes that organizations use these data in attempts to sway audiences.

Table 4. Frequency of data used for the work of crediting or discrediting different types of entities by organization political leaning.

	Organization		Individual		Issue		Group		Total	
	Cons	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons	Lib
Crediting	84	317	47	120	148	106	27	44	306	587
Discrediting	57	12	822	90	144	8	110	17	1133	127
Neutral	13	6	6	1	41	10	10	5	70	22
Totals	154	335	875	211	333	124	147	66	1509	736

Organizations' affective rhetorical work most often focuses on individuals (48% of total instances of data), who are more commonly discredited (41% of total instances of data) rather than credited (7%). Groups are also more often discredited (6%) than credited (3%). Specific organizations (22%), however, are more likely to be credited (18%) than discredited (3%), with some referenced neutrally, as well (.01%). Issues (19%) are also more likely to be credited (11%) than discredited (7%) and are referenced with the highest amount of neutral data (.02%) in the corpus.

4.2.1 Partisan Crediting & Discrediting. Conservative and liberal organizations use data affectively in very different ways. Conservative organizations use data for discrediting 75% of the time, for crediting 20% of the time, and neutrally 5% of the time. Contrastingly, the liberal organizations use data for crediting 79% of the time, for discrediting 17% of the time, and neutrally 4% of the time (Table 4). Freedom Partners, the conservative energy organization, presents data almost exclusively for discrediting (FP: 98%). When analyzing how data are used by these groups to credit or discredit over the one month duration of the corpus, the conservative organizations in our corpus relied on using data to discredit (their primary affective strategy) entities, particularly earlier on in the month. As Election Day drew nearer, the conservative groups discredited with data less frequently. In contrast, the liberal organizations overall use data to credit entities (their primary affective strategy) with increasing frequency working up to Election Day and, particularly, as reactions to the election results were posted—despite the loss of the presidential race. These data-crediting reactions focused in particular on the results of congressional and state level races, and many aim to celebrate the wins and highlight the organization's role in making that happen.

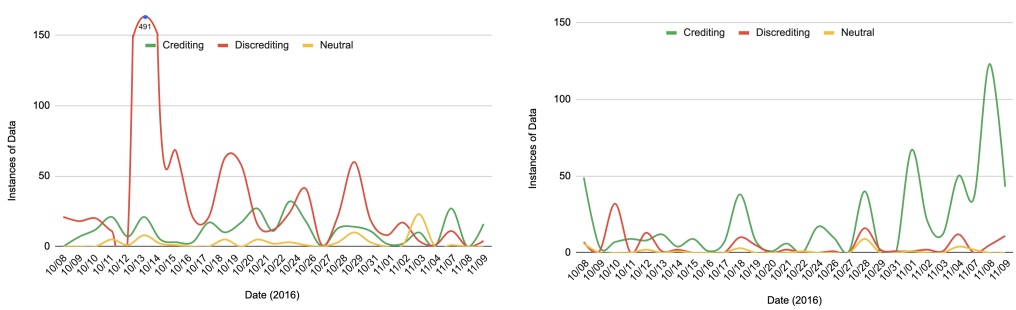

Fig. 5. Distribution of Conservative (Left) and Liberal (Right) Crediting, Discrediting, and Neutral Data over Time

4.2.2 Entities of Affective Focus. Each of the liberal organizations in our corpus primarily referenced *organizations* when presenting data (335 instances; 46% of liberal data use). The liberals in this corpus

credited organizations in 43% of their total data use (317 instances). This is often self-congratulatory in nature, such as, "Today, EMILY's List, the nation's *largest* resource for women in politics, congratulated Val Demings on being elected to represent Florida's Tenth Congressional District" (EMILY, 08 November). In contrast, the conservatives in this corpus only reference organizations in 10% of the data they present, and discredit them slightly more often than they credit them.

Conservative organizations use data most frequently to reference *individuals* (875 instances; 58% of conservative data), almost always discrediting them (54% of total conservative references). For example: "Under Obama, 23.7 million Americans between 25 and 54 years old aren't working. Home ownership fell to 62%, the lowest rate in 51 years" (RNC, 10 October). In striking contrast, only 29% of liberal data references individuals and only 12% of liberal data discredits individuals (Table 4). When data reference individuals or groups, however, all organizations are more likely to discredit than to credit. Finally, three of the four conservative single-issue organizations in our corpus presented data about their issue area more frequently than their liberal counterparts. The contrast between how liberal and conservative organizations use data rhetorically points to more than a difference in style. The liberal organizations' abundant use of data to credit themselves and other ally entities demonstrates their priority to use data as opportunities to bolster the legitimacy of their organization and policy positions. By using data to discredit opposing entities, the conservative organizations call the legitimacy of others into question.

4.3 Evaluating Logos: Are the Data Empirical?

The conservative and liberal organizations use the same proportions of both empirical (conservatives: 85%; liberals: 85%) and unempirical data. While there is a significant prevalence of empirical data, in the month prior to the 2016 election, increases in unempirical data correspond to increases in empirical data, as well. We also note that there is not an increase in unempirical data right before Election Day.

4.3.1 Un-Empirical Data.

Projections. In this corpus, 7% of data (164 instances) are projections (6% of conservative data (97 instances); 9% of liberal data (69 instances)). Sometimes projected data are based on robust empirical data, but are still speculative, such as: "The proposed expansion...is touted to add 1,000 good-paying jobs over five years" (EMILY, 10 October). However, projected data also sometimes rely on more dubious quantitative grounding. For example, by treating votes for past proposed policies as current reality, one post quoted news articles from decades ago to discredit modern politicians: "The [1999 bill] would reduce taxes by $782 billion over 10 years by reducing income tax rates by 10 percent" (FP, 14 October). FP's use of evidence was unempirical, as they omitted further explanation of the observed results and instead presented then-hypothetical implications as evidence to discredit a politician who was allied with the lawmaker who supported this bill.

Extrapolations. In some instances, data appear to be derived from valid empirical data, extending quantitative assumptions to generate new data. Such instances (3% of corpus; 76 instances) are coded as extrapolated data and account for 4% of all conservative data (60 instances), and only 1% of liberal data (11 instances). For example, "…nearly 300 little boys and girls are walked to their death every single day. That's nearly 600 parents, 1200 grandparents, countless siblings" (NRL, 12 October). Here, the initial 300 instances are categorized as empirical data, however, the subsequent "600 parents, 1200 grandparents, and countless siblings" are each unempirical extrapolations.

Rhetorical Fabrications. Finally, some unempirical data (103 instances; 5% of all data) are instances of data presented in the context of logical fallacies (e.g. false causality, ad hominem, slippery slope, tautologies) and other linguistic turns that change the intended meaning of the data. We refer to

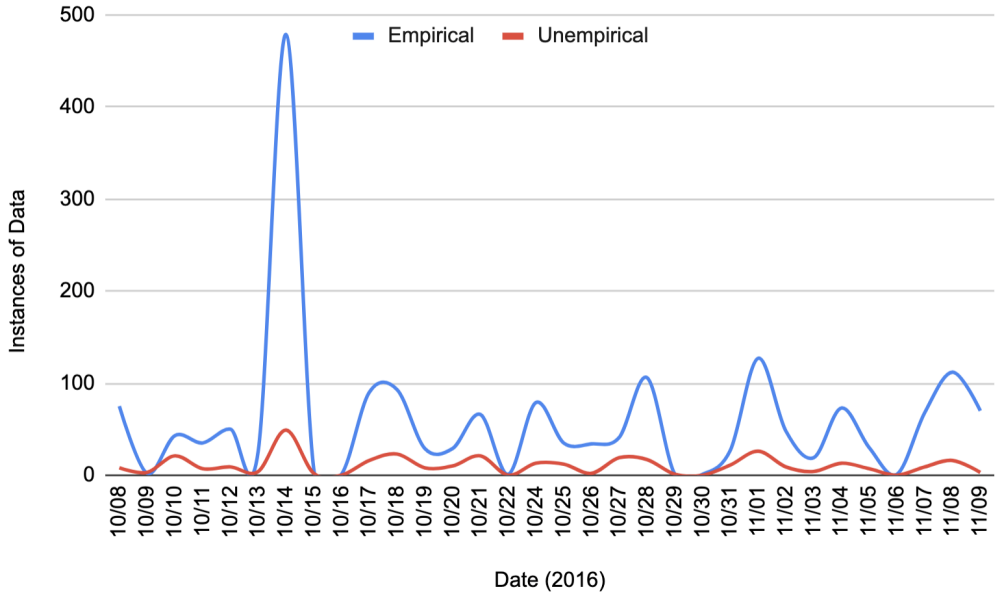

Fig. 6. Distribution of empirical and unempirical posts over time

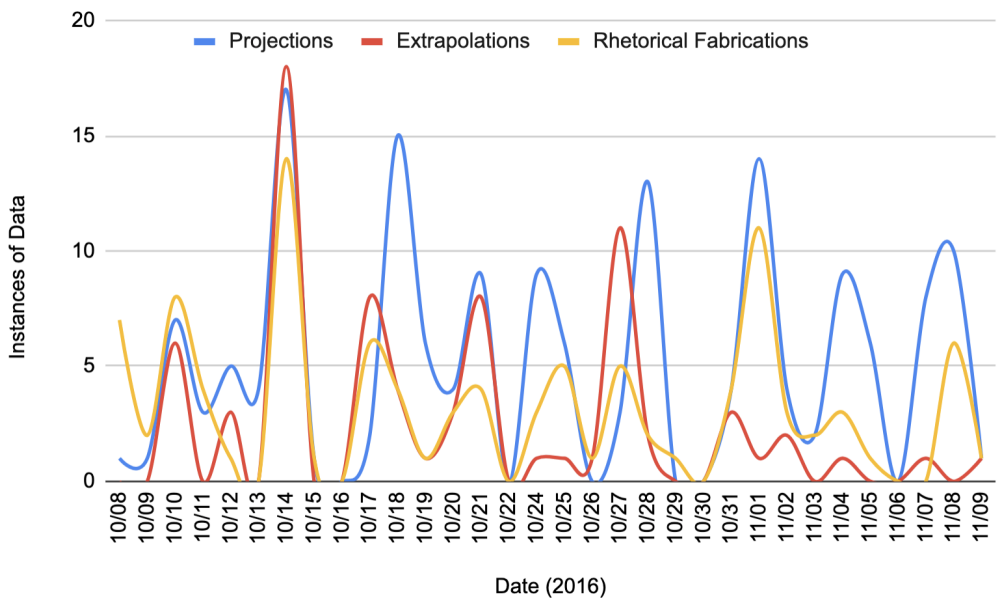

Fig. 7. Distribution of unempirical data over time

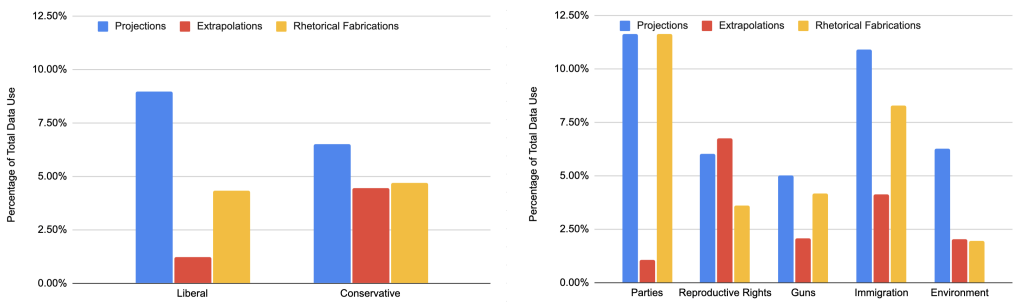

Fig. 8. Normalized aggregate use of different types of unempirical data by ideology (Left) and Issue Area (Right)

such instances as rhetorically fabricated. This type of data use appears in 5% of all conservative data (74 instances) and 4% of liberal data (32 instances). One common example of rhetorically fabricated data relies on the logical fallacy of false causality. For example, the conservative energy group writes: "Despite Russ Feingold's claims that he is a deficit hawk, the national debt tripled during his 18 years in the Senate" (FP, 14 October). Here, the organization implies a false causality between a politician's tenure in office and the national debt within the same sentence.

Other logical fallacies used to present data include ad hominem attacks, such as the DNC's statement, "The Donald Trump we saw on stage tonight is the same Trump who has spent the last 30 years demeaning women, avoiding taxes, and using dangerous and divisive rhetoric" (DNC, 10 October). Instead of discussing the debate and the candidate performances, the DNC draws from "30 years" of behavior as evidence to discredit him. Other organizations use vague references to data that lead to inductive hyperbole, where they draw a conclusion stronger than what their evidence supports, for example: "In addition to ethnic food and cultural activities, increasing numbers of immigrants have also contributed to the spread of many behaviors and practices contrary to American values. While the food can be delicious, the "new ideas" are often malicious" (FAIR, 11 October).

4.3.2 Partisan Un-Empirical Data Use. When analyzing the distribution of unempirical data over the course of the month preceding the 2016 elections (Figure 7), we note that as elections drew nearer, projections were the primary form of unsound data used, whereas prior to 18 October, extrapolations and rhetorical fabrications are used more frequently. While liberal and conservative organizations use similar rates of unempirical data, the liberal organizations rely mostly on *projected* data while conservative organizations use the three types of unempirical data more evenly.

4.3.3 Notable Issue-Specific Behavior. Each of the five organizational pairs used different distributions of unempirical data, but the biggest difference among the pairs was between the political parties and the issue organizations. The political parties used rhetorically-fabricated data more than single-issue advocacy organizations and rarely used extrapolated data. The single issue advocacy organizations all used projected data most frequently, with different issue areas using rhetorically fabricated with varied frequency. It may be, then, that different ideological and issue audiences (Figure 8) have different norms or appetites for rhetorically-fabricated data.

Table 5. Key contrasts between liberal and conservative uses of data.

Conservative Organizations	**Liberal Organizations**	**Similarities**
Post a larger volume of text, both with slightly higher frequency (206 posts), greater word count (168,202 words), lower frequency of data (every 111 words)	Post a smaller volume of text, both with slightly lower frequency (131 posts), smaller word count (65,905 words), higher frequency of data (every 89 words)	
Primarily count money (45% of all conservative instances of data use)	Primarily count people (47% of all liberal instances of data use)	
Use specific numbers 53% of the time (799 instances) Use narrative descriptions 7% of the time (101 instances)	Use specific numbers 39% of the time (286 instances) Use narrative descriptions 22% of the time (162 instances)	Use specific numbers more than other forms of specificity Use similar rates of vague data (C: 284 instances; 19% L: 153 instances; 21%) Use similar rates of calculations (C: 335 instances; 22% L: 140 instances; 19%)
Use data primarily to discredit (1,133 instances; 75%) Reference individuals the most (875 instances; 58%); of those instances, 96% are used to discredit (822 instances)	Use data primarily to credit (587 instances; 80%) Reference organizations the most (335 instances; 46%); of those instances, 94% are used to credit (317 instances)	All organizations discredit individuals more frequently than other entities (Total: 912 instances; 41% C: 822 instances; 73% L: 90 instances; 12%)
When conservatives use unempirical data (236 instances), it is more evenly distributed across data types: Projected: 7% (98 instances) Extrapolated: 4% (67 instances) Rhetorical Fabrications: 5% (71 instances)	When liberals use unempirical data (107 instances), it is predominantly projected: Projected: 9% (66 instances) Extrapolated: 1% (9 instances) Rhetorical Fabrications: 4% (32 instances)	Use similar rates of both empirical data... (C: 1,274 instances; 84% L: 624 instances; 85%) ...and unempirical data (C: 236 instances; 16% L: 107 instances; 15%)

5 DISCUSSION

Just as prior research has found that conservative and liberal entities use language differently [115, 124], our research has found that conservatives and liberals also use *data* differently (Table 5).

Conservative organizations' data rhetoric reflects the conservative worldview that imagines the role of government as that of a strict father — emphasizing law, self-reliance and self-discipline [68]. The conservative organizations in our corpus post more instances of data in the context of more frequent and lengthy posts (mirroring this worldview's emphasis on teaching through language), predominantly count money (reflecting the importance of fiscal self-reliance), use more specific

numbers, use data predominantly to discredit individuals (mirroring a disciplinary use of language), and are more varied in their use of unempirical data.

In contrast, liberal organizations' data rhetoric often reflects the liberal worldview that imagines the role of government as that of a nurturing parent, emphasizing empathy and helping others [68]. The liberal organizations in our corpus post fewer instances of data in fewer and shorter posts (mirroring this worldview's emphasis on teaching through action rather than language), primarily count people (which aligns with their operationalization of justice as caring for people within the broader society), use data predominantly to credit organizations (mirroring this worldview's emphasis on taking moral actions that cultivate happiness), and primarily use projections when sharing unempirical data. Although our qualitative analysis exists within a specific moment in time—when the presidency had been under Democrat power and the senate and house majorities were both held by Republicans—worldviews are remarkably stable, neurologically-instantiated structures [68] As such, we would expect the rhetoric from both conservative and liberal worldviews to exhibit some degree of constancy across communicative contexts, whether posts are about elections or issues happening at the federal, state, or local level. The degree of transferability across contexts would, however, be a compelling direction for future research.

While there are significant differences between how liberal or conservative organizations use data rhetoric, organizations across the ideological spectrum are using data rhetoric in ways that align with their worldviews to help make sense of reality and imagine their political futures (see also [24, 67, 74]). Imaginaries are "collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order" [60]. Löfflmann, for example, describes how a Trump-aligned PAC leverages rhetoric to evoke a populist security imaginary, seeking to portray ideologically opposed people as threats to security [74]. Our research extends this line of scholarship to characterize how ideologically diverse organizations are using data-as-rhetoric to evoke political imaginaries.

Across the ideological spectrum, Browne and Diehl characterize a transformation in the political imaginary:

> *There are new forms of political experiences, online and offline movements, and a new kind of political consciousness, which does not necessarily follow the logic of political institutions and is sometimes anti-political or post-truth. These phenomena are signs of a deep transformation of the political imaginary [31].*

While the modern *political imaginary* does not necessarily follow logic and has moved beyond 'truth,' insofar as facts do not matter as much as the futures that are characterized and imagined through data rhetoric (e.g., [30, 74]), the modern *sociotechnical imaginary* of the data economy considers data to be a valuable and essential commodity. Under this sociotechnical imaginary, data is inherently valuable and, as such, must be protected [101].

5.1 Trajectories for Data Rhetoric in The Sociotechnical Imaginary of the Data Economy

These paradoxical views toward data present an inflection point for sociotechnical scholars and our trajectories for future research. We may, on one hand, align our research to support the imaginary of the data economy, leading to a focus on maintaining or salvaging the 'inherent' value of facts. Here we see a trajectory for future research that builds off of existing efforts to identify and minimize disinformation, focusing on the problems of "incorrect knowledge" [15] and "innocent readership" [58]. Existing strategies to interrogate the use of data in the post-truth era tend to center around fact-checking degrees of the validity of information, in inferring authorial intention around the use of false data (i.e., distinguishing between misinformation and disinformation), and exploring other consequences of misrepresented facts. These perspectives lead researchers to ask

questions like whether or not people accept 'facts' (e.g., [32]) or how 'lies' propagate algorithmically (e.g., [121]).

"Correcting" misinformation has had mixed results, with some scholars suggesting that these strategies can actually backfire [125], as the presentation of these evaluations are imbued in politics of their own. Many audiences distrust fact-checks, and question the political leanings and ulterior intentions of fact checkers (e.g. [99]) Many visible actors (e.g., advocacy organizations [57], law makers [41], journalists [28, 81], and platforms [1]) engage in fact-checking the fact-check, presenting further contextual distinctions and nuance to argue that a given fact-check is erroneous. Waisbord attributes a lack of common epistemology to the competing perspectives on "truth-telling anchored in different premises" [123]. Although many fact checking websites include longform explanations, their reductionist grading or labeling "only amplify the binary and make truth the purview of gatekeepers, intermediaries, and validators" [99].

Further complicating research along this trajectory, Calo et al. note that mis/disinformation campaigns often "involve true information and reasonable opinion" [34, p. 1]. In our corpus, $\tilde{1}$5% of the data are unempirical and only 5% are erroneously misleading. As such, researchers will likely need to scale up the kind of analysis that we have undertaken here, analyzing how data is "cooked" beyond its degree of validity. Our analytic framework provides initial structure for the development of natural language processing models for detecting instances of quantitative data and evaluating their logical/rhetorical validity. While our coded corpus can serve as an initial training set, our experience qualitatively coding these texts suggests a few challenges and opportunities for future research:

- **Detecting individual instances of data.** The largest set of discrepancies in inter-rater reliability in this research were in how the two coders bracketed each instance of quantitative data. Models will need to be able to identify quantitative data that is non-numerical and narratively descriptive. For example, '2016 is on track to be the warmest year on record' does include an instance of quantitative data, but the instance occurs on the words 'warmest year' and not '2016.' Additionally, models will need to be able to discriminate between quantitative data used for counting and quantitative data used as context (e.g., the date in the previous example).
- **Scaffolding assessments of the empiricism of data**. Some researchers may be inclined to develop models with the capacity to algorithmically identify and flag instances of unempirical data or rhetorically fabricated data. But, we see value in using computation alongside human critical thinking—both to render judgment about unempirical data as well as to help engender information literacy among audiences. However, the design challenges in doing so are great: How does one best scaffold audiences in understanding what may be problematic in text (i.e., more than just issuing a generic warning along with a link to an authoritative source on the topic)? How might that scaffolding need to vary by audience, issue or platform? How does one convey degrees of certainty about the output of the algorithm? What are optimal degrees of transparency in different contexts? And how do we design for information literacy about data rhetoric practices that might be emotionally manipulative but not egregiously false? Design research for enabling the assessment of logos in order to foster critical thinking about data rhetoric may present some of the greatest challenges for future mis/disinformation research.

Although the data economy may have had its roots in newspaper publishing and advertising [55], the rhetoric (and infrastructure) of the modern data economy resonates with the culture (and

expertise) of Silicon Valley, where Ferrari argues that technology is constructed through mainstream discourses that are both populist and technocratic [47]. That the solutions explored for mis/disinformation have been predominantly technical, then, befits this imaginary.

5.2 Trajectories for Data Rhetoric in The Political Imaginary

This research highlights a second imaginary — one that isn't focused on assessing and preserving the presumably-inherent value of data and truth. The political imaginary, in contrast, sidesteps logic and truth, privileging the communication and persuasion of the political future it sees and is invested in [31]. If whether or not a fact is empirical ceases to matter, this alternate imaginary suggests a trajectory for future research in which we are called to change our analytical focus from fact to feeling, exploring the affective work of data, instead (see also [56]. The political imaginary sees the value of data in their rhetorical ability to persuade and emotionally manipulate [56], what Boler and Davis refer to as the "weaponization" of information on digital media [24]. There are many research trajectories in this space for sociotechnical scholars, as well:

- **Detecting the crediting and discrediting work of data.** In our Twitter corpus, data was leveraged to conduct affective work that primarily credited or discredited entities. Identifying and highlighting this work might serve to mitigate its negative influences [23]. Scaling up this analysis would likely require additional NLP research. The challenge here is that the logical argument surrounding an instance of data can span across multiple sentences (which is the unit of analysis used in existing NLP research on propaganda detection [91]). The entity that data are connected to can also be the issue area itself, implicit in the text, making detection highly contextual and domain dependent. Further, many statements may initially appear as neutral, such as the example, "he'll sign all pro-life bills" (EMILY, 10 October). Yet here, EMILY is using one individual's willingness to sign bills in order to discredit that individual.
- **Detecting other affective work of data.** While this research has focused on distinguishing between positive and negative affect (crediting and discrediting), affective work takes on more nuanced forms that might be explored. Data rhetoric might be employed to incite anger or fear [51, 65, 113], two of the more common emotions studied in political communication, though some scholars have argued for more research that explores positive emotional manipulation such as hope and enthusiasm [107]. Applying NLP techniques to this more nuanced task of detecting other affective work may also help in the development of predictive models to alert audiences to the goals of such affective manipulation. Boler and Davis, for example, suggest that new media audiences are manipulated in order to solicit GOTV efforts, donations, volunteers, and for assistance in the propagation of information [24]. Chadwick's analysis of the repertoires of action of single-issue advocacy organizations provides an alternate possible framework for this analysis [35]. Future research would be well suited to explore the connections between the more nuanced affective work of data rhetoric and its correlation to these genres of solicitations: What are affective data trying to compel audiences to do?
- **Modeling the information propagation of data rhetoric.** Just as Adamic and Glance found that conservative blogs were more likely to interlink to one another to bolster their arguments [11], future work should explore how—and what kinds of —data rhetoric are propagated through and between different online spaces. This trajectory for future research will be particularly important as more political advocacy organizations start to employ data rhetoric in shorter and more highly-networked forms of social media communication. In this evolving ecosystem of networked communication, we have the opportunity to uncover the extent to which different forms of data rhetoric (e.g., more specific data vs. more narrative data or data with more positive affect vs. more negative affect) are propagated to new social

contexts via shares or retweets and which are preserved within their original social context (e.g., via likes or comments). This direction for future research would enable more precise modeling of different social structures of data rhetoric, whether echo chambers, epistemic bubbles [89], or something entirely new.

- **Exploring interactions among additional data imaginaries.** Ferrari characterized the sociotechnical imaginary of Silicon Valley as a populist technocracy [47]. As this imaginary is embodied through platform design. In 2017, Facebook, for example, comes to portray itself as a "*social infrastructure*" for civic engagement "uniquely suited…to address global issues, such as climate change and terrorism" [47]. How, then, do the UX design features and/or the algorithmic curation of this third imaginary interact with these other political and sociotechnical imaginaries? As user behavior changes and/or as public outcry becomes unavoidable, how do platforms adapt? And how do these adaptations, in turn, affect these other imaginaries? For example, following the 2016 elections, many sociotechnical spaces adopted new design features, such as fact check ratings [1] or sensitive content warnings [4] in response to the use of the platform by end users enacting a political imaginary. Yet these design responses do not engage with the political imaginary; they, instead, adapt their own imaginaries in response. Understanding the push and pull of imaginaries through design, the ways the values and motivations of imaginaries interact versus move in response to but siloed from each other.

6 CONCLUSION

Our research explores data rhetoric, assessing how advocacy organizations use quantitative data in their social media outreach. By applying our analytic framework to the blog posts of 10 political advocacy organizations during the month preceding the 2016 elections, we highlighted the ways in which political advocacy organizations leverage data in affective politics and political propaganda. First, in the most literal sense, the *ethos* of data are ***counting something***, employing a variety of genres of presentation. Second, these instances of data often appear alongside *pathos*, which do the work of ***crediting or discrediting some entity***. Third, recognizing the soundness of the logical argument related to the instance of data and the entity that it is crediting or discrediting reveals its *logos*, which is the relative ***soundness of how organizations use data for their advocacy work***.

Our research demonstrates that data are used in highly situated and rhetorically sophisticated ways. Ideologically aligned organizations wield facts though language in similar ways, likely contributing to "echo-chambers" [89] in which audiences are not merely exposed to expressions of similar values and ideas, but are also acclimated to reading similar styles of data rhetoric. While this research suggests that data rhetoric practices may be shaped by ideological worldview, future research would be well served to better understand whether or to what extent the echo-chambers of the data rhetoric also serve to reciprocally shape or even legitimize certain worldviews in return.

Our comparative analysis of the data rhetoric of conservative and liberal organizations also highlights a number of different rhetorical strategies in use across political ideologies, a better understanding of which should enable both individuals and organizations to tailor their data use to different audiences—or perhaps even find ways to reach out across our political, linguistic, and data rhetoric differences in more empathic ways.

This research has several limitations. First, we sampled organizations that spent the most money influencing the 2016 US election cycle. Future research studying a broader sample of organizations—including less well-funded organizations or advocacy organizations whose missions center around less ideologically divisive issues, for example—might highlight other features of the rhetorical work of data. Additionally, our sample only included two organizations in each issue-category.

This left us unable to make strong claims about similarity in data rhetoric within a given issue. Future research would be well served to focus on a broader set of organizations in each issue area or policy field in order to see more robust results regarding issue-focused data practices. Further, the distribution of posts among the different advocacy organizations affect results, especially as this analysis focused on only ten organizations. Despite normalizing data frequency for each organization, some organizational practices were inevitably more analytically influential than others. Our analysis notes where this is the case and also makes note of organizational outliers, but future research with a broader sample of organizations within a policy field would also help to mitigate this limitation. Understanding the data rhetoric of organizations outside of the US–especially because of the globalization of political polarization [10]—is also critical to developing a more generalizable understanding of electoral data rhetoric. Finally, the ecosystem of social media platforms is dynamic and the ways in which advocacy organizations use those platforms has changed since 2016. As such, future research should also explore the ways in which data rhetoric is used across a broader range of platforms with different affordances for communication modality and audience engagement across other elections. These contexts would make a compelling site to apply or extend our analytic framework.

While post-truth rhetoric focuses our attention on the ways in which people's worldviews are shaped and, especially, reinforced by the information curated in social media; here, we find that our worldviews may also inform and shape the ways in which we use data rhetoric. Understanding the ways that quantitative data has been "cooked" [49] through language across both the social media ecosystem and the ideological spectrum is essential to increasing the reflexivity of our collective data rhetoric practices.

REFERENCES

[1] [n. d.]. About fact-checking on Facebook and Instagram. https://www.facebook.com/business/help/2593586717571940
[2] [n. d.]. Election results, 2016. https://ballotpedia.org/Election_results,_2016#State_executive_offices
[3] [n. d.]. Gubernatorial elections, 2016. https://ballotpedia.org/Gubernatorial_elections,_2016
[4] [n. d.]. Reporting and viewing sensitive media | twitter help. https://help.twitter.com/en/safety-and-security/sensitive-media
[5] 2016. Election 2016 - ballot measures. *CNN* (Nov 2016). https://www.cnn.com/election/2016/results/ballot-measures
[6] 2016. Federally Focused 527s by Sector. https://www.opensecrets.org/527s/527cmtes.php?level=S&cycle=2016
[7] 2016. Ideology; Single Issue Sector Background. https://www.opensecrets.org/industries/indus.php?ind=Q
[8] 2016. Ideology; Single Issue Sector Summary. https://www.opensecrets.org/industries/indus.php?ind=Q
[9] 2016. Top voting issues in 2016 election. https://www.pewresearch.org/politics/2016/07/07/4-top-voting-issues-in-2016-election/
[10] Alan I. Abramowitz and Steven Webster. 2016. The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Studies* 41 (2016), 12–22. https://doi.org/10.1016/j.electstud.2015.11.001
[11] Lada A Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*. 36–43.
[12] Saud Alashri, Srinivasa Srivatsav Kandala, Vikash Bajaj, Roopek Ravi, Kendra L Smith, and Kevin C Desouza. 2016. An analysis of sentiments on Facebook during the 2016 US presidential election. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 795–802.
[13] Adriana Alvarado Garcia and Christopher A Le Dantec. 2018. Quotidian Report: Grassroots Data Practices to Address Public Safety. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–18.
[14] Adriana Alvarado Garcia, Alyson L Young, and Lynn Dombrowski. 2017. On making data actionable: How activists use imperfect data to foster social change for human rights violations in Mexico. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–19.
[15] PEN America. 2017. Faking news: Fraudulent news and the fight for truth. *New York, NY: Author* (2017).
[16] Claudia Aradau and Tobias Blanke. 2017. Politics of prediction: Security and the time/space of governmentality in the age of big data. *European Journal of Social Theory* 20, 3 (2017), 373–391.
[17] Bruno Araújo and Hélder Prior. 2020. Framing Political Populism: The Role of Media in Framing the Election of Jair Bolsonaro. *Journalism Practice* (2020), 1–17.

[18] Matt Bai. 2005. The framing wars. *The New York Times Magazine* 17 (2005), 38–45.
[19] Vian Bakir and Andrew McStay. 2018. Fake news and the economy of emotions: Problems, causes, solutions. *Digital journalism* 6, 2 (2018), 154–175.
[20] Vian Bakir and Andrew McStay. 2020. Empathic Media, Emotional AI, and the Optimization of Disinformation. In *Affective Politics of Digital Media*. Routledge, 263–279.
[21] Pablo Barbera. 2018. Explaining the spread of misinformation on social media: Evidence from the 2016 US presidential election. In *Symposium: Fake News and the PoliticsofMisinformation. APSA*.
[22] Eric PS Baumer, Francesca Polletta, Nicole Pierski, and Geri K Gay. 2017. A simple intervention to reduce framing effects in perceptions of global climate change. *Environmental Communication* 11, 3 (2017), 289–310.
[23] Eric PS Baumer, Jordan Sinclair, and Bill Tomlinson. 2010. America is like Metamucil: fostering critical and creative thinking about metaphor in political blogs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1437–1446.
[24] Megan Boler and Elizabeth Davis. 2020. *Affective politics of digital media*. Routledge.
[25] Chris Bopp, Ellie Harmon, and Amy Voida. 2017. Disempowered by data: Nonprofits, social enterprises, and the consequences of data-driven work. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3608–3619.
[26] Geoffrey C Bowker. 2000. Biodiversity datadiversity. *Social studies of science* 30, 5 (2000), 643–683.
[27] Danah Boyd and Jayshree Sarathy. 2022. Differential Perspectives: Epistemic Disconnects Surrounding the US Census Bureau's Use of Differential Privacy. *Harvard Data Science Review (Forthcoming)* (2022).
[28] Petter Bae Brandtzaeg, Asbjørn Følstad, and Maria Ángeles Chaparro Domínguez. 2018. How journalists and social media users perceive online fact-checking and verification services. *Journalism practice* 12, 9 (2018), 1109–1129.
[29] Nadia M Brashier and Daniel L Schacter. 2020. Aging in an era of fake news. *Current directions in psychological science* 29, 3 (2020), 316–323.
[30] Craig Browne. 2019. The modern political imaginary and the problem of hierarchy. *Social Epistemology* 33, 5 (2019), 398–409.
[31] Craig Browne and Paula Diehl. 2019. Conceptualising the political imaginary: an introduction to the special issue. , 393–397 pages.
[32] Katie Byrd and Richard S John. 2021. Lies, damned lies, and social media following extreme events. *Risk analysis* (2021).
[33] Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The guardian* 17 (2018), 22.
[34] Ryan Calo, Chris Coward, Emma S Spiro, Kate Starbird, and Jevin D West. 2021. How do you solve a problem like misinformation? *Science advances* 7, 50 (2021), eabn0481.
[35] Andrew Chadwick. 2007. Digital network repertoires and organizational hybridity. *Political Communication* 24, 3 (2007), 283–301.
[36] Pamela Johnston Conover, Virginia Gray, and Steven Coombs. 1982. Single-issue voting: Elite-mass linkages. *Political Behavior* 4, 4 (1982), 309–331.
[37] Martin De Saulles. 2015. *Information 2.0: New models of information production, distribution and consumption*. facet publishing.
[38] Jeffrey Dowd. 2017. Racial Discourse and Partisan Blogs: How Online Commenters Manage the Partisan Divide. *Sociological Perspectives* 60, 1 (2017), 33–51.
[39] Robert Draper. 2013. Inside the Power of the NRA. *New York Times* 15 (2013).
[40] Y Roselyn Du, Lingzi Zhu, and Benjamin KL Cheng. 2019. Are numbers not trusted in a "post-truth" era? An Experiment on the Impact of Data on News Credibility. *Electronic News* 13, 4 (2019), 179–195.
[41] Paul Egan. 2021. GOP lawmakers file bills that would require 'Fact checkers' to register with the State. https://www.freep.com/story/news/politics/2021/05/11/michigan-gop-lawmakers-want-fact-checkers-register-face-fines/5045565001/
[42] Kyle Endres and Kristin J Kelly. 2018. Does microtargeting matter? Campaign contact strategies and young voters. *Journal of Elections, Public Opinion and Parties* 28, 1 (2018), 1–18.
[43] Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication* 43, 4 (1993), 51–58.
[44] Sheena Erete, Emily Ryou, Geoff Smith, Khristina Marie Fassett, and Sarah Duda. 2016. Storytelling with data: Examining the use of data by non-profit organizations. In *Proceedings of the 19th ACM conference on Computer-Supported cooperative work & social computing*. 1273–1283.
[45] Robert Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. *Berkman Klein Center Research Publication* 6 (2017).

[46] Henry Farrell and Daniel W Drezner. 2008. The power and politics of blogs. *Public choice* 134, 1 (2008), 15–30.
[47] Elisabetta Ferrari. 2020. Technocracy meets populism: The dominant technological imaginary of Silicon Valley. *Communication, Culture & Critique* 13, 1 (2020), 121–124.
[48] Michael Gamon, Sumit Basu, Dmitriy Belenko, Danyel Fisher, Matthew Hurst, and Arnd Christian König. 2008. BLEWS: Using blogs to provide context for news articles.. In *ICWSM*. 60–67.
[49] Lisa Gitelman, Virginia Jackson, Daniel Rosenberg, Travis D Williams, Kevin R Brine, Mary Poovey, Matthew Stanley, Ellen Gruber Garvey, Markus Krajewski, Rita Raley, et al. 2013. Data flakes: An afterword to "raw data" is an oxymoron. (2013).
[50] Erving Goffman. 1959. The moral career of the mental patient. *Psychiatry* 22, 2 (1959), 123–142.
[51] David J Grüning and Thomas W Schubert. 2022. Emotional Campaigning in Politics: Being Moved and Anger in Political Ads Motivate to Support Candidate and Party. *Frontiers in Psychology* 12 (2022).
[52] Lauren Guggenheim, S Mo Jang, Soo Young Bae, and W Russell Neuman. 2015. The dynamics of issue frame competition in traditional and social media. *The ANNALS of the American Academy of Political and Social Science* 659, 1 (2015), 207–224.
[53] Ian Hacking, Tim Hacking, et al. 1990. *The taming of chance*. Number 17. Cambridge University Press.
[54] Eszter Hargittai and Aaron Shaw. 2013. Digitally savvy citizenship: The role of internet skills and engagement in young adults' political participation around the 2008 presidential election. *Journal of Broadcasting & Electronic Media* 57, 2 (2013), 115–134.
[55] Timandra Harkness. 2021. The history of the data economy: Part I: The birth of customer insight. *Significance* 18, 6 (2021), 12–15.
[56] Sun-ha Hong. 2020. Fuck Your Feelings: The Affective Weaponisation of Facts and Reason. In *Affective Politics of Digital Media: Propaganda by Other Means*, Megan Boler and Elizabeth Davis (Eds.). Routledge, New York, NY, USA, 86–100. https://doi.org/10.4324/9781003052272
[57] Mike Howell. 2022. Fact check: Fact-checkers falsely claim they are fact-checkers. https://www.heritage.org/progressivism/commentary/fact-check-fact-checkers-falsely-claim-they-are-fact-checkers
[58] Elle Hunt. 2016. What is fake news? How to spot it and what you can do to stop it. *The Guardian* 17 (2016), 15–16.
[59] S Mo Jang and P Sol Hart. 2015. Polarized frames on "climate change" and "global warming" across countries and states: Evidence from Twitter big data. *Global Environmental Change* 32 (2015), 11–17.
[60] Sheila Jasanoff and Sang-Hyun Kim. 2015. *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. University of Chicago Press.
[61] Vera Khovanskaya and Phoebe Sengers. 2019. Data Rhetoric and Uneasy Alliances: Data Advocacy in US Labor History. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 1391–1403.
[62] Andrei P Kirilenko and Svetlana O Stepchenkova. 2014. Public microblogging on climate change: One year of Twitter worldwide. *Global environmental change* 26 (2014), 171–182.
[63] Klaus Krippendorff. 2018. *Content analysis: An introduction to its methodology*. Sage publications.
[64] Udo Kuckartz and Stefan Rädiker. 2019. Analyzing intercoder agreement. In *Analyzing Qualitative Data with MAXQDA*. Springer, 267–282.
[65] Adi Kuntsman. 2020. Reverberation, Affect, and Digital Politics of Responsibility. In *Affective Politics of Digital Media*. Routledge, 69–85.
[66] Matthew J Lacombe. 2019. The political weaponization of gun owners: The national rifle association's cultivation, dissemination, and use of a group social identity. *The Journal of Politics* 81, 4 (2019), 1342–1356.
[67] George Lakoff. 2006. *Thinking points: Communicating our American values and vision*. Macmillan.
[68] George Lakoff. 2010. *Moral politics: How liberals and conservatives think*. University of Chicago Press.
[69] Martha Lampland. 2010. False numbers as formalizing practices. *Social Studies of Science* 40, 3 (2010), 377–404.
[70] Richard R Lau and Gerald M Pomper. 2001. Negative campaigning by US Senate candidates. *Party Politics* 7, 1 (2001), 69–87.
[71] Richard R Lau, Lee Sigelman, Caroline Heldman, and Paul Babbitt. 1999. The effects of negative political advertisements: A meta-analytic assessment. *American Political Science Review* (1999), 851–875.
[72] Eric Lawrence, John Sides, and Henry Farrell. 2010. Self-segregation or deliberation? Blog readership, participation, and polarization in American politics. *Perspectives on Politics* 8, 1 (2010), 141–157.
[73] Dave Levitan. 2015. Unspinning the Planned Parenthood Video. https://www.factcheck.org/2015/07/unspinning-the-planned-parenthood-video/
[74] Georg Löfflmann. 2022. 'Enemies of the people': Donald Trump and the security imaginary of America First. *The British Journal of Politics and International Relations* 24, 3 (2022), 543–560.
[75] David Lyon. 2019. Surveillance capitalism, surveillance culture and data politics. *Data Politics: Worlds, Subjects, Rights. Abingdon: Routledge* (2019), 64–77.

[76] Catherine MacPhail, Nomhle Khoza, Laurie Abler, and Meghna Ranganathan. 2016. Process guidelines for establishing intercoder reliability in qualitative studies. *Qualitative research* 16, 2 (2016), 198–212.
[77] David Malakoff. 2017. A matter of fact.
[78] Lev Manovich. 2011. Trending: The promises and the challenges of big social data. *Debates in the digital humanities* 2, 1 (2011), 460–475.
[79] Carol Mason. 2018. *Killing for Life.* Cornell University Press.
[80] Amaryllis Mavragani and Konstantinos P Tsagarakis. 2019. Predicting referendum results in the Big Data Era. *Journal of Big Data* 6, 1 (2019), 1–20.
[81] Megan McArdle. 2016. Fact-checking's infinite-regress problem. https://www.bloomberg.com/opinion/articles/2016-12-23/fact-checking-s-infinite-regress-problem?leadSource=uverify+wall
[82] Nolan McCarty and Eric Schickler. 2018. On the theory of parties. *Annual Review of Political Science* 21 (2018), 175–193.
[83] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.
[84] Lee McIntyre. 2018. *Post-truth.* MIt Press.
[85] Laura McKenna and Antoinette Pole. 2004. Do blogs matter? Weblogs in American politics. American Political Science Association.
[86] Yelena Mejova, Venkata Rama Kiran Garimella, Ingmar Weber, and Michael C Dougal. 2014. Giving is caring: understanding donation behavior through email. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 1297–1307.
[87] Isaac Nahon-Serfaty. 2018. *Strategic communication and deformative transparency: Persuasion in politics, propaganda, and public health.* Routledge.
[88] Alessandro Nai. 2020. Going negative, worldwide: Towards a general understanding of determinants and targets of negative campaigning. *Government and Opposition* 55, 3 (2020), 430–455.
[89] C Thi Nguyen. 2020. Echo chambers and epistemic bubbles. (2020).
[90] Jonathan A Obar, Paul Zube, and Clifford Lampe. 2012. Advocacy 2.0: An analysis of how advocacy groups in the United States perceive and use social media as tools for facilitating civic engagement and collective action. *Journal of information policy* 2 (2012), 1–25.
[91] Vitaliia-Anna Oliinyk, Victoria Vysotska, Yevhen Burov, Khrystyna Mykich, and Vítor Basto Fernandes. 2020. Propaganda Detection in Text Data Based on NLP and Machine Learning.. In *MoMLeT+ DS*. 132–144.
[92] Brian L Ott. 2017. The age of Twitter: Donald J. Trump and the politics of debasement. *Critical studies in media communication* 34, 1 (2017), 59–68.
[93] Gordon Pennycook and David G Rand. 2020. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of personality* 88, 2 (2020), 185–200.
[94] Andrea Pereira, Elizabeth Ann Harris, and Jay J Van Bavel. 2018. Identity concerns drive belief: The impact of partisan identity on the belief and dissemination of true and false news. (2018).
[95] Kathleen H Pine and Max Liboiron. 2015. The politics of measurement and action. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3147–3156.
[96] Peter Pomerantsev and Michael Weiss. 2014. *The menace of unreality: How the Kremlin weaponizes information, culture and money.* Vol. 14. Institute of Modern Russia New York.
[97] Vincent Raynauld and André Turcotte. 2018. "Different strokes for different folks": Implications of voter microtargeting and appeal in the age of Donald Trump. In *Political marketing in the 2016 US presidential election*. Springer, 11–28.
[98] Tyler T Reny, Loren Collingwood, and Ali A Valenzuela. 2019. Vote switching in the 2016 election: How racial and immigration attitudes, not economics, explain shifts in white voting. *Public Opinion Quarterly* 83, 1 (2019), 91–113.
[99] Evelyn Ruppert. 2018. *Sociotechnical imaginaries of different data futures.*
[100] Lydia Saad. 2016. Party groups agree on importance of big election issues. https://news.gallup.com/poll/192353/party-groups-agree-import-big-election-issues.aspx
[101] Jathan Sadowski. 2019. When data is capital: Datafication, accumulation, and extraction. *Big data & society* 6, 1 (2019), 2053951718820549.
[102] Dietram A Scheufele. 1999. Framing as a theory of media effects. *Journal of communication* 49, 1 (1999), 103–122.
[103] Jonathon P Schuldt, Sara H Konrath, and Norbert Schwarz. 2011. "Global warming" or "climate change"? Whether the planet is warming depends on question wording. *Public opinion quarterly* 75, 1 (2011), 115–124.
[104] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1139–1148.

[105] Subir Sinha. 2017. Fragile hegemony: Modi, social media and competitive electoral populism in India. *International Journal of Communication* 11, 2017 (2017), 4158–4180.
[106] Alina Sîrbu, Dino Pedreschi, Fosca Giannotti, and János Kertész. 2019. Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PloS one* 14, 3 (2019), e0213246.
[107] Cigdem V Sirin and José D Villalobos. 2019. The study of discrete emotions in politics. In *Oxford Research Encyclopedia of Politics*.
[108] Theda Skocpol and Alexander Hertel-Fernandez. 2016. The Koch network and republican party extremism. *Perspectives on Politics* 14, 3 (2016), 681–699.
[109] Dominic Spohr. 2017. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review* 34, 3 (2017), 150–160.
[110] Kate Starbird. 2017. Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 11.
[111] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.
[112] Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, and Robert M Mason. 2014. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. *IConference 2014 Proceedings* (2014).
[113] Luke Stark. 2020. Empires of feeling: social media and emotive politics. In *Affective politics of digital media*. Routledge, 298–313.
[114] Charles D Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. 2016. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories. (2016).
[115] Karolina Sylwester and Matthew Purver. 2015. Twitter language use reflects psychological differences between democrats and republicans. *PloS one* 10, 9 (2015), e0137422.
[116] Zeynep Tufekci. 2018. How social media took us from Tahrir Square to Donald Trump. *MIT Technology Review* 14 (2018), 18.
[117] Daphne Van der Pas. 2014. Making hay while the sun shines: Do parties only respond to media attention when the framing is right? *The International Journal of Press/Politics* 19, 1 (2014), 42–65.
[118] Chris J Vargo, Lei Guo, and Michelle A Amazeen. 2018. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New media & society* 20, 5 (2018), 2028–2049.
[119] Maya Vassileva. 2019. Fake News as a "Rhetorical Device". *Rhetoric and Communications Journal* (2019).
[120] Amy Voida, Ellie Harmon, Willa Weller, Aubrey Thornsbury, Ariana Casale, Samuel Vance, Forrest Adams, Zach Hoffman, Alex Schmidt, Kevin Grimley, et al. 2017. Competing currencies: Designing for politics in units of measurement. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 847–860.
[121] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
[122] Howard Wainer. 2013. *Visual revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot*. Psychology Press.
[123] Silvio Waisbord. 2018. Truth is what happens to news: On journalism, fake news, and post-truth. *Journalism studies* 19, 13 (2018), 1866–1878.
[124] Sean P Wojcik, Arpine Hovasapian, Jesse Graham, Matt Motyl, and Peter H Ditto. 2015. Conservatives report, but liberals display, greater happiness. *Science* 347, 6227 (2015), 1243–1246.
[125] Thomas Wood and Ethan Porter. 2019. The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior* 41, 1 (2019), 135–163.
[126] Julia K Woolley, Anthony M Limperos, and Mary Beth Oliver. 2013. The 2008 presidential election, 2.0: A content analysis of user-generated political Facebook groups. In *New Media, Campaigning and the 2008 Facebook Election*. Routledge, 87–108.