Received: 1 May 2023 | Revised: 19 December 2023 | Accepted: 20 December 2023

DOI: 10.1002/asi.24869

RESEARCH ARTICLE

JASIST WILEY

What's in your PIE? Understanding the contents of personalized information environments with PIEGraph

Deen Freelon[1,2] | Meredith L. Pruden[3] | Daniel Malmer[2,4] | Qunfang Wu[5] | Yiping Xia[6] | Daniel Johnson[2,4] | Emily Chen[2] | Andrew Crist[2]

[1]Annenberg School for Communication, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[2]Center for Information, Technology, and Public Life, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[3]School of Communication and Media, Kennesaw State University, Kennesaw, Georgia, USA

[4]Hussman School of Journalism and Media, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[5]Inequality in America Initiative, Harvard University, Cambridge, Massachusetts, USA

[6]Department of Communication & Journalism, Texas A&M University, College Station, Texas, USA

Correspondence

Deen Freelon, Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA 19104, USA.

Email: dfreelon@upenn.edu

Funding information

John S. and James L. Knight Foundation, Grant/Award Number: G-2019-58787; William and Flora Hewlett Foundation, Grant/Award Number: 2019-9047

Abstract

Social media have long been studied from *platform-centric* perspectives, which entail sampling messages based on criteria such as keywords and specific accounts. In contrast, *user-centric* approaches attempt to reconstruct the personalized information environments users create for themselves. Most user-centric studies analyze what users have accessed directly through browsers (e.g., through clicks) rather than what they may have seen in their social media feeds. This study introduces a data collection system of our own design called PIEGraph that links survey data with posts collected from participants' personalized X (formerly known as Twitter) timelines. Thus, in contrast with previous research, our data include much more than what users decide to click on. We measure the total amount of data in our participants' respective feeds and conduct descriptive and inferential analyses of three other quantities of interest: political content, ideological skew, and fact quality ratings. Our results are relevant to ongoing debates about digital echo chambers, misinformation, and conspiracy theories; and our general methodological approach could be applied to social media beyond X/Twitter contingent on data availability.

1 | INTRODUCTION

As digital communication technologies have diffused throughout society, scholarly attention has focused on the contents and influence of what we call personalized information environments (PIEs). This term refers to automatically populated digital information spaces assembled through expressed user preferences and/or algorithmic predictions: examples include social media feeds, RSS feeds, lists of subscribed podcasts, and similar applications. As such environments have become part of everyday life for many, concerns have grown about the nature and quality of the information that appears within them. One research tradition investigates the extent to which personalized environments function as *echo chambers*: self-reinforcing information feeds that screen out ideologically unfriendly content (Pariser, 2011; Sunstein, 2018). Such research usually attempts to ascertain the ideological balance of political content in the data, with imbalances tilting too far to one side qualifying as echo chambers. A more recent research direction involves normative questions of information health that distinguish–implicitly or explicitly–between "good" and "bad" types of information. Since the US 2016 presidential election, much attention has focused on the bad, under conceptual banners including *misinformation*, *disinformation*, *problematic context*, *fake news*, *information pollution*, and many others. Such research has undertaken a number of goals, three of the most prominent being defining the problem (Freelon & Wells, 2020; Wardle & Derakhshan, 2017), mapping its scope (Allcott & Gentzkow, 2017; Farhall et al., 2019), and testing the efficacy of proposed solutions (Walter et al., 2020). Personalized information environments have attracted attention across disciplinary lines, with scholars from information science, computer science, communication, political science, public policy, and public health making substantial contributions.

Our study emerges from a methodological tradition that attempts to understand information health by mapping the contents of participants' personalized information environments. This *user-centric* approach departs sharply from the *platform-centric* approach that dominates the social scientific study of social media (Breuer et al., 2022; Christner et al., 2022; Ohme et al., 2023). Studies of the latter type typically proceed by collecting and analyzing data from one or more social media platforms based on sampling criteria such as keywords, account names, time periods, and/or geographic locations. The platform-centric approach is effective at such tasks as identifying influencers, determining the prevalence of the phenomena of interest, and finding predictors of user engagement, but it is not ideal for understanding users' personalized information environments.

In contrast, user-centric approaches attempt to reconstruct personalized information environments using methods that require more user interaction than simply extracting trace data from a social media platform. Using a combination of surveys and tools that collect data from individual participants' devices or accounts, such methods can link opinions and attitudes with media content that users consume or opt in to. User-centric methods have grown in popularity in recent years, spurred by the availability of commercial providers of such data (e.g., Comscore and Netquest) and research software development by academics (Menchen-Trevino, 2016; Ohme et al., 2023; Reeves et al., 2021). We developed PIEGraph, the software system powering the current research, to collect user-centric data that is inaccessible to existing applications. While most user-centric research examines the websites or apps participants visit by clicking links or entering web addresses directly, PIEGraph reconstructs each user's chronological X/Twitter timeline based on the unique set of users they follow. Within that platform, we argue ours is a more faithful representation of users' information environments as it incorporates content participants have chosen to view, not the small minority of content they have time to click on.

In this study, we use PIEGraph to conduct an initial investigation into the personalized information environments of 790 study participants. We are principally concerned with four distinct quantities: (1) the total amount of content, (2) the amount of content classified as "political," (3) the environments' ideological leanings, and (4) their levels of factual quality. We analyze these quantities cross-sectionally, longitudinally, and model them in regression analyses, with some of our findings being consistent with prior research and others departing from it.

2 | HOW PERSONALIZED INFORMATION ENVIRONMENTS HAVE BEEN MEASURED

The concept of the "information ecosystem," to which this special issue is devoted, encompasses components that govern the production, distribution, and consumption of such media, including laws, policies, markets, technological design choices, psychological biases, and more (Carter et al., 2023; Norris & Suomela, 2017). This study investigates one particular part of the information ecosystem, which we call the *personalized information environment*, which consists of collections of media units (posts) assembled by an algorithm and/or the user's choices of which accounts to follow (cf "content streams," Bayer et al., 2020; Carter et al., 2023; "curated flows," Thorson & Wells, 2016). Such environments are one of the main ways users interact with social media—it is the default view for Facebook, X/Twitter, Instagram, TikTok, and other platforms. PIEs determine what users see and do not see within a platform—they are the individualized universes within which users decide how to direct their attention.

In the fields of communication and political science, the standard method of studying personalized information environments has been the survey questionnaire. Relevant questions vary in their granularity: some ask about broad media types (e.g., radio, TV, newspapers, social media, etc.), while others offer preset or freeform options to indicate interest in specific outlets. Such questions can only capture a general sense of how people allocate their attention to media—even when they ask users to state which outlets or sites they visit most often, they can reveal at most a minuscule sliver of their respondents' media diets. Social media platforms present an especially vexing challenge to this method: knowing that Facebook is the United States' most frequently mentioned "top site" (Knight Foundation, 2022) tells us little about the content users see when they log on. The topical composition, political leaning, and amount of misinformation (among many other variables) of a user's friend or follower network may vary widely depending on the accounts therein and the individual-level account settings the user chooses.

Information science scholars have long used the diary method to investigate people's information behaviors. It entails instructing participants to manually record details of their media consumption activities on a regular basis using "diaries," which can incorporate multiple media types. Carter and Mankoff (2005) wrote that "[t]he diary study is a method of understanding participant behavior and intent in situ," in which "participants control the timing and means of capture" (p. 900, emphasis original). Their participants used a number of audio, visual, and tangible objects as components of their media diaries. Carter and Mankoff noted a need for "situated annotation of captured event[s]" (p. 908) through specialized tools. More recent studies have used online tools and mobile applications to meet this need. Rieh et al. (2010) collected over 2400 diary entries from 333 respondents to study people's credibility assessment of everyday information on the web, while Narayan et al. (2011) examined the phenomenon of information avoidance in everyday settings by qualitatively analyzing diary entries totaling 468 participant days.

Scholars from a variety of disciplines have used diary methods to study the consumption of political content online. Saltz et al. (2021) combined in-depth interviews, diary, and co-design methods to investigate Americans' perceptions of their information environments and their attitudes toward platforms' misinformation labels. Moe and Ytre-Arne (2022) argued that diaries enable a kind of "investigation beyond the moment, and beyond single platforms or providers" (p. 44). For example, Gulyas et al. (2019) used diaries to study how audiences consume local news online, whereas Mihelj et al. (2022) combined interviews with media diaries to study audience engagement with COVID-19 news. The use of diaries goes beyond learning about news consumption patterns, with Beckers et al. (2021) focusing on political learning through the news in their diary study.

The diary method's major limitations stem from human fallibility: it tends to under-capture events that are infrequent or, alternatively, frequent but last only briefly. Further, because of the labor involved, long-term diary studies tend to have high attrition rates (Vandewater & Lee, 2009). And a recent systematic review found that self-reported media use correlates only moderately with automatically logged traces of digital activity (Parry et al., 2021). Various user-centric methods of information environment research have emerged in recent years to compensate for these limitations.

3 | USER-CENTRIC RESEARCH ON SOCIAL MEDIA ENVIRONMENTS

Since its inception, the study of social media has been dominated by platform-centric approaches. Such studies collect data matching certain publicly available criteria, usually one or more keywords or authorship by particular sets of users (e.g., all posts within a given timeframe authored by members of Congress). While platform-centric research has been effective in answering a broad range of research questions, it can tell us little about what ordinary users encounter when browsing social media. Christner et al. (2022) define user-centric methods as those that "trac[e] the comprehensive media usage of an individual on the client side" (p. 80). In contrast with platform-centric methods, user-centric approaches attempt to see as the user sees at greater resolutions than are possible with surveys or diaries. Another difference is that, since participants are essentially granting permission for their data to be used in research, the researchers typically obtain informed consent, which is infeasible for large-N platform-centric studies. User-centric methods may use existing social media APIs, as ours does; scraping techniques that mine cached browser data (Haim & Nienierza, 2019; Menchen-Trevino, 2016); or download whatever appears on participants' screens at regular intervals (e.g., Reeves et al., 2021). We might use a platform-centric approach if we are concerned with how much misinformation or debunked content exists on social media (Farhall et al., 2019; Ng & Loke, 2021). But if we want to know how many people are exposed to or share misinformation, we will need to use data on what content users have selected or viewed—that is, a user-centric approach (Grinberg et al., 2019; Guess et al., 2019; Osmundsen et al., 2021). We argue that user-centric perspectives are more applicable to questions of information quality because misinformation that no one consumes is of little relevance.

User-centric research on personalized information environments has addressed several key questions and topics. One of the most widely studied areas incorporates a cluster of theories including selective exposure, political polarization, and ideological echo chambers. These kinds of studies generally attempt to measure how politically diverse participants' information environments are, based on the longstanding normative fear that people will use digital affordances to screen out ideologically challenging content (see, e.g., Sunstein, 2018). Findings reveal such echo chambers to be fairly rare–most participants have ideologically diverse news preferences (Bentley et al., 2019; Bruns, 2019; Guess, 2021; Nelson & Webster, 2017). Moreover, news constitutes a relatively small proportion of most people's information diets (Cronin et al., 2022; Stier et al., 2022; Wojcieszak et al., 2022). Research in the COVID-19 era has found that the pandemic has increased attention to the news, as people seek trustworthy scientific, medical, and policy information to help structure their daily lives (Altay et al., 2022; Nelson & Lewis, 2022). A related collection of studies has focused on the transition of news audiences from desktop to mobile devices (Fedeli & Matsa, 2018; Nelson & Lei, 2018), with mobile news access being associated with increased ideological diversity (Yang et al., 2020).

Misinformation has been another major focus of user-centric research since 2016, when the topic surged onto the social science agenda (Grinberg et al., 2019; Kucharski, 2016; Williamson, 2016). Existing studies have found misinformation to be rare overall, but concentrated among political conservatives and older adults. In particular, such individuals have been observed to consume and share misinformation at higher rates than others (Allen et al., 2020; Grinberg et al., 2019; Guess et al., 2019). Misinformation shared on social media is likely to be consistent with whatever ideology the sharing user holds (Guess et al., 2021). Heavily polarized individuals—those who report strong dislike of political outgroup members—have been observed to share misinformation at high rates, with this effect being much stronger for Republicans than Democrats (Osmundsen et al., 2021). The audiences of sites dedicated to producing false “clickbait” content have been characterized as small, disloyal, and chronically online (Nelson & Taneja, 2018). Conservatives have been shown to be more susceptible to online misinformation specifically concerning COVID-19 (Calvillo et al., 2020). But, perhaps counterintuitively, visiting low-credibility sites is not consistently related to holding false beliefs (Weeks et al., 2021).

4 | OUR USER-CENTRIC APPROACH: PIEGRAPH

Existing user-centric approaches break down into several categories, each with its own advantages and disadvantages. Many user-centric researchers purchase data from commercial brokers such as Comscore and NetQuest, which reveal the hyperlinks participants click on. This approach is effective in discovering how participants focus on content of high interest but omits content they may have seen in their social media feeds but not clicked on. Understanding such content is important, especially given the influence of headlines as sources of information about the world beyond direct experience (Geer & Kahn, 1993; Janét et al., 2022). Another technique, called data donation, relies on participants' generosity in exporting their own social media data using tools provided by the platform (Araujo et al., 2022; Halavais, 2019; Ohme et al., 2023; Pfiffner & Friemel, 2023). But such donations typically include only messages the participants create themselves rather than those posted by accounts they follow. The Screenomics project is unique in using client-side software to take screenshots of participants' mobile and desktop devices every 5 seconds (Reeves et al., 2021). While this powerful technique can collect content from any app (including offline ones), the process of converting data-rich pages into flat images strips away much content of empirical value, including hyperlinks and image URLs.

To our knowledge, the approach we adopt with our PIEGraph system is unique in the universe of user-centric methods. It uses an opt-in X/Twitter API endpoint (*statuses/home_timeline*) that returns all content in our participants' feeds—that is, messages by users they follow. Thus, we can reconstruct each of our participants' personalized X/Twitter environments for the entire study period, with no clicks or screen time required. Our system's major contribution is to capture complete, user-constructed environments of X/Twitter for analysis, an accomplishment unprecedented at the current scale as far as we are aware. One major difference between PIEGraph and other user-centric methods is that we have no way of knowing how much of their timeline content our participants have actually viewed. However, we argue that the content that populates X/Twitter users' timelines represents the kinds of content our participants are generally interested in. In other words, we assume that participants who have followed (for example) many conservative/right-leaning accounts are probably interested in conservative content regardless of how often they actually use X/Twitter. Our empirical results support this assumption.

5 | RESEARCH QUESTIONS

Our research questions offer an initial exploration into the data collected through PIEGraph, drawing on existing user-centric research on news, politics, and misinformation. We begin with basic questions about the amount of data across all PIEs (RQ1a and b), proceeding to the respective distributions across participants (RQ2a–c), change over time (RQ3a–c), and survey-derived predictors (RQ4a–c) for political content, ideological skew, and fact quality. Coincidentally, our dataset covers the periods of time preceding and following Elon Musk's purchase of X/Twitter, which allows us to analyze the extent to which our participants' PIEs changed in the Musk era.

- RQ1a: What is the total number of messages collected across all personalized information environments?
- RQ1b: How does the number of messages change over time?
- RQ2a: What are the amount and distribution of political content across all our sample's users?
- RQ2b: What are the mean ideological skew and distribution of political content across all our sample's users?
- RQ2c: What are the mean factual reporting skew and distribution of political content across all our sample's users?
- RQ3a: How does political content volume change over time?
- RQ3b: How does mean ideological skew change over time?
- RQ3c: How does mean fact quality change over time?
- RQ4a: Which survey characteristics predict individual political content volume?
- RQ4b: Which survey characteristics predict individual ideological skew?
- RQ4c: Which survey characteristics predict individual factual reporting skew?

6 | DATA AND METHODS

We collected our data using a software system of our own design called PIEGraph, which differs from the other user-centric systems described above both technically and conceptually. We began by contracting with the research firm Dynata to supply a panel of participants demographically matched to the US Census (see SI section B in Data S1). Each participant was required to hold an active X/Twitter account to be eligible for the study. They each filled out a survey containing a battery of items inquiring about their demographic details, ideological and partisan commitments, media consumption preferences, and beliefs in various conspiracy theories (see SI section C in Data S1). They were then offered the opportunity to grant read-only access to their X/Twitter timelines for the research team. Participants were informed that their X/Twitter data would be collected and analyzed anonymously. Participants remained anonymous to us throughout the entire research process, as Dynata handled the compensation logistics and generated unique alphanumeric IDs for each participant. Of the 19,021 people who took our survey, 1020 (5.4%) chose to offer us access to their X/Twitter timelines. Participants were solicited between December 2021 and May 2022, and data collection occurred between December 2021 and April 2023 and concluded on April 4, 2023, when PIEGraph's access to the X/Twitter API was revoked.

PIEGraph collected and stored complete timeline data for each of these participants (i.e., posts by users they follow). Data were collected through X/Twitter's v2 API, making the process completely invisible to participants. Instructions for how to remove oneself from the project were posted to the project web site, and 94 participants did so, leaving 926 valid participants at the time of data analysis. Each participant's X/Twitter data was stored alongside their Dynata ID so that we could link it with their survey data. Participants who completed the survey and granted access to their X/Twitter timelines were compensated $5.

We were especially interested in the hyperlinks populating our participants' personalized information environments, so we drew on data from the Media Bias Fact Check (MBFC; https://mediabiasfactcheck.com/) project, which measures the ideological leaning and fact quality of over 3400 English-language news and political sites. MBFC measures both ideology and factual reporting on unidimensional scales, with the former ranging from extreme left to extreme right, and the latter ranging from "Very High" to "Very Low."[1] Its factual reporting scores represent subjective evaluations of the sites it indexes. To attain the highest score, a site must fulfill the following criteria: "always [be] factual, sources to credible information, and makes immediate corrections to incorrect information, and has never failed a fact check in either news reporting or op-eds" (Media Bias Fact Check, 2020, n.p.). A site that "rarely uses credible sources and is not trustworthy for reliable information at all" (Media Bias Fact Check, 2019, n.p.) would receive MBFC's lowest factual reporting score. The site's scores correlate highly with alternative domain-based information quality metrics (Lin et al., 2022) and have been used in a number of studies (Aires et al., 2019; Chen et al., 2022; Ribeiro et al., 2018). See SI section A in Data S1 for additional details on how we collected and preprocessed the raw MBFC data.

After our X/Twitter data had been collected, we used a Python script to link each participant's timeline posts with their survey data using their Dynata IDs. We excluded all participants whose posts contained fewer than 20 MBFC-classified link instances (so called because

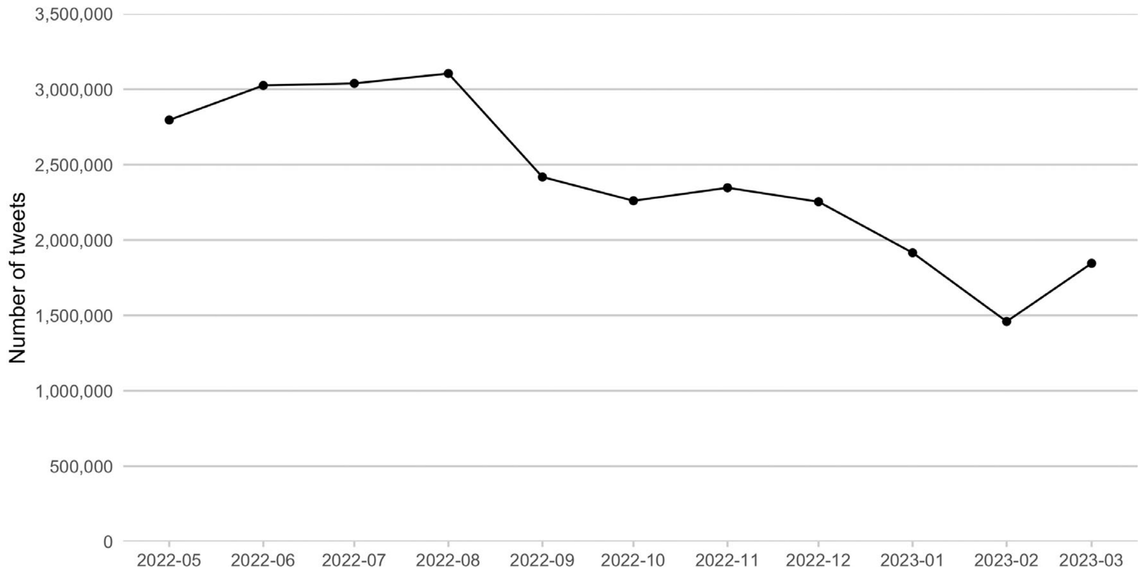

FIGURE 1 Total unique posts over time.

multiple instances of the same link counted toward the total), leaving us with a final sample size of $n = 790$ participants. For the link instances that appeared in these participants' PIEs, we calculated mean ideology and factual reporting scores that we use as outcome variables in the statistical models reported below.

7 | RESULTS

To answer RQ1a, we tallied the total number of unique posts across all PIEs, which came to 106,929,041.[2] Individual participants' PIE sizes varied widely—from a minimum of 0 posts on some days to a maximum of 1,981 posts on a single day—depending primarily on how many users each participant followed.[3] Figure 1 displays the total number of posts per month starting in May 2022, the month participant recruitment concluded, and March 2023, the final complete month of data collection. A general decline begins from August to September, tapering off temporarily but resuming from November through February with a brief rally from February to March. Addressing RQ1b, this analysis provides some evidence that our participants' PIEs have decreased in size since Elon Musk took control of X/Twitter at the end of October 2022. This could be the result of accounts closing, users going dormant on the platform and/or trying out platform alternatives like Mastodon, or X/Twitter's spam/bot removal efforts. Other relevant factors may include Twitter filing suit against Musk over the sale in July, as well as the leaked announcement around the company's issues with information security in August. While it is outside the scope of this study to determine which of these variables caused the decline in unique post volume, it is likely some combination thereof.

To answer our cluster of research questions about the respective quantities and distributions of political, ideological, and fact quality content (RQ2a–c), we begin by operationalizing "political content" as domains classified by MBFC. This rests on the assumption that the organization has attained comprehensive coverage of most news and political sites that are widely known in the United States. Of all unique posts, 11,251,711 (10.5%) contained at least one MBFC-classified link. Figure 2 shows the distribution of MBFC classification proportions across participants, with participants' percentile positions comprising the x-axis. In other words, the chart represents two different proportions: that of MBFC-classified links out of all of each participant's PIE links (y-axis), and that of each participant ranked in descending order of their corresponding y-axis proportion (x-axis). It reveals, for example, that participants at the 90th percentile are seeing PIEs in which 25% of links are classified as "political" by our definition, and everyone below that rank is seeing <25% political links. Such content drops very quickly—to about 50% by the 98th percentile, to 13% by the 75th, to 3%–4% by the 50th percentile (median). Moving to ideological leaning (RQ2b), Figure 3 displays participant percentiles by the mean ideology scores of their MBFC-classified links. It reveals that our participants' timelines are mostly left-leaning, with roughly 20% having a right-of-center ideology mean, and the remaining 80% falling to the left of center. Finally, Figure 4 plots the

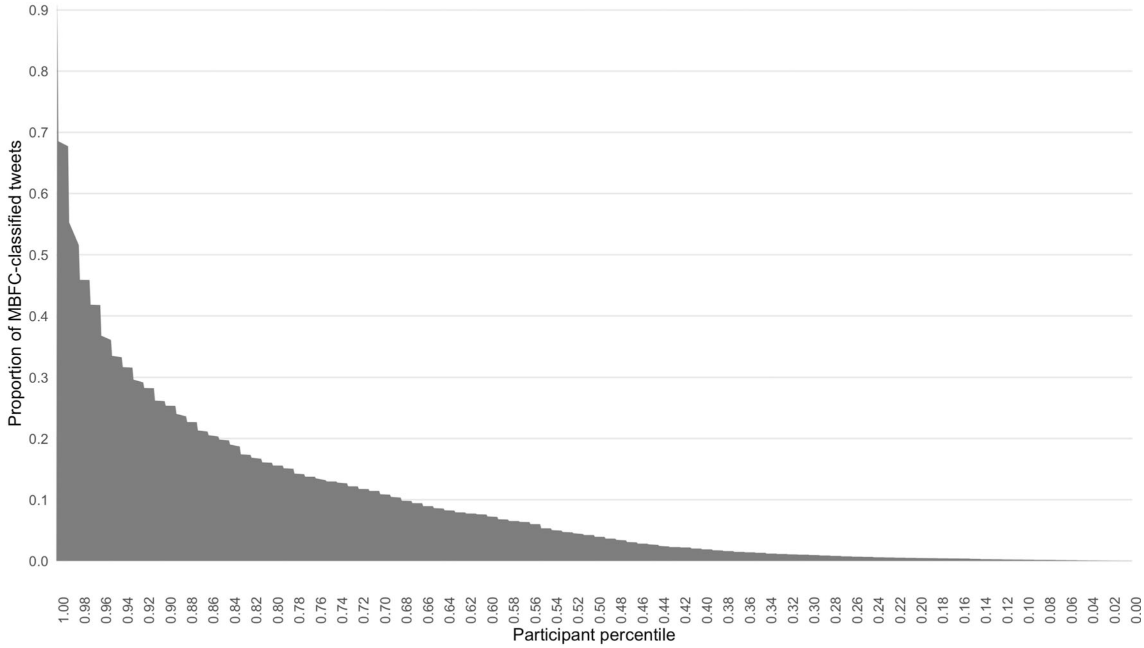

FIGURE 2 Distribution of political posts by participant percentile.

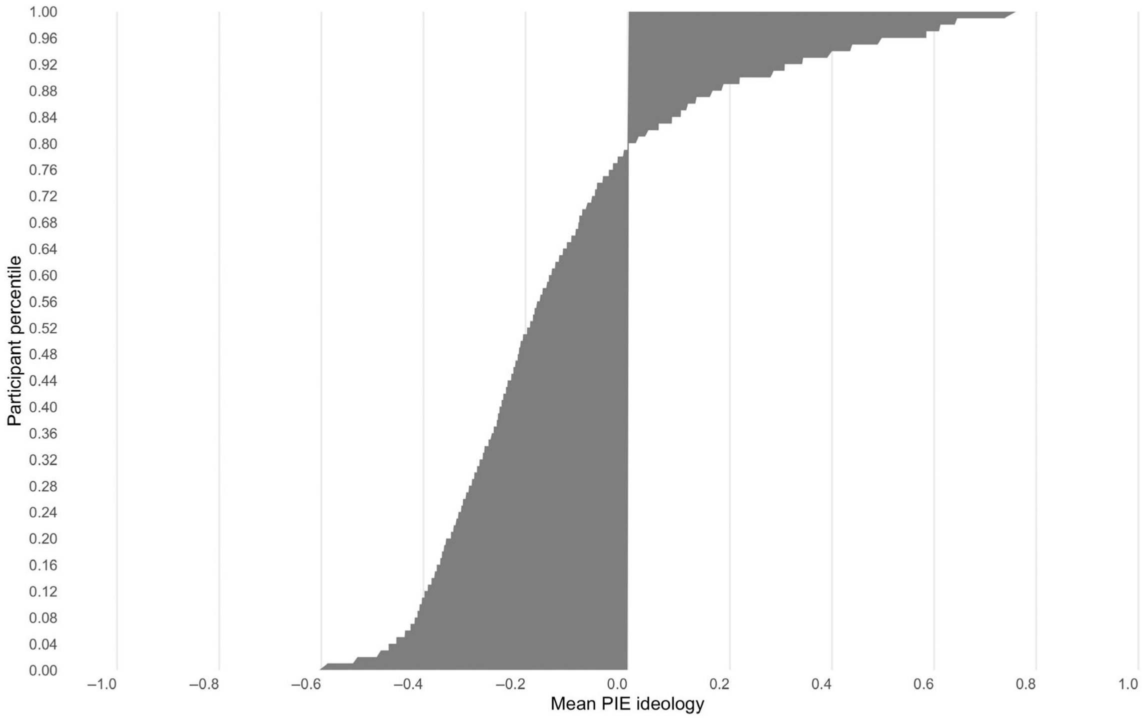

FIGURE 3 Distribution of mean ideology scores by participant percentile.

participant percentile against the individual mean fact quality of PIE links (RQ2c). Again, higher *y*-axis values correspond to greater degrees of mean fact quality. Encouragingly, only participants in the bottom 3% see mean fact quality scores at or near the lowest two levels, Low and Very Low (corresponding to *y*-axis values of 0.2 and 0, respectively). About 90% of participants see mean fact quality values between 0.4 (Mixed) and 0.8 (High), while the top 6% of PIEs feature mean scores between "High" and "Very High" (scored as 1).

RQ3a–c relate to how our three quantities of interest change over time. Figure 5 answers RQ3a, plotting the average monthly proportions of posts containing at least

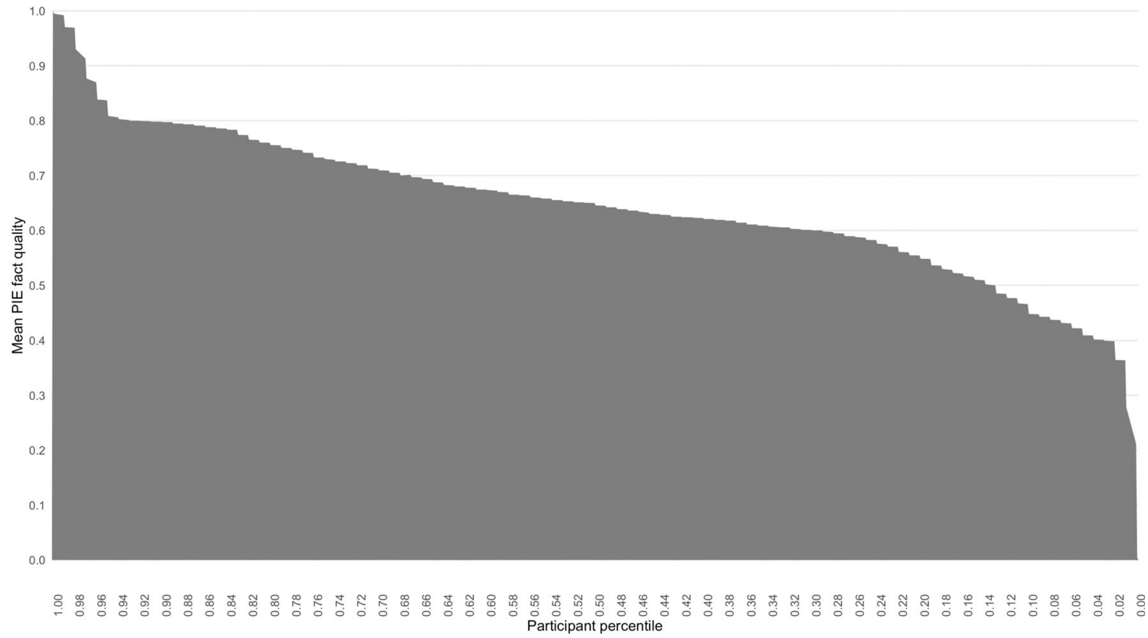

FIGURE 4 Distribution of mean fact quality scores by participant percentile.

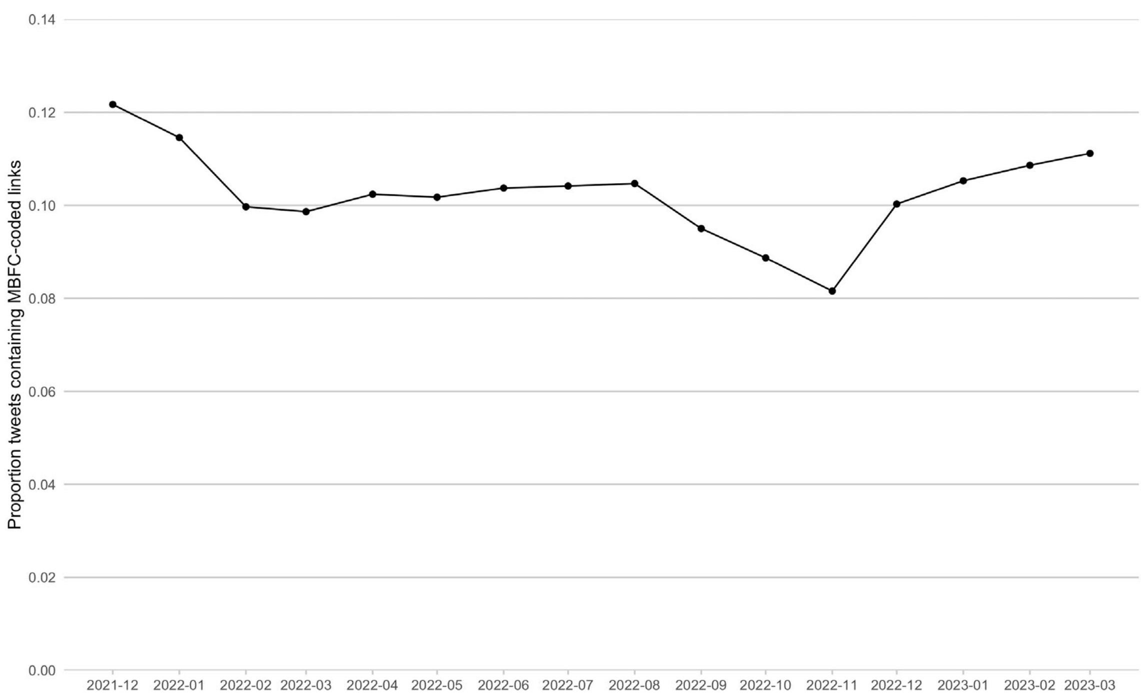

FIGURE 5 Average monthly proportions of political posts.

one MBFC-coded link out of all posts. Overall, the degree of longitudinal change is quite limited apart from a slight decrease between August and November. Otherwise, the proportions tend to hover close to 0.1. Monthly mean ideological leaning across all PIEs (Figure 6; RQ3b) shows a similar pattern, with values remaining slightly below zero (left of center) after recruitment concludes in May. The decreasing mean ideological leaning over time reflects the fact that more of our left-leaning participants entered our sample later during the recruitment process. Similarly, in Figure 7 we see that mean fact quality stabilizes after May slightly above the "Mostly Factual" level (0.6), addressing RQ3c.

Figure 8a–c answers RQ4a–c, which inquires about the extent to which individual-level variables can predict PIE content. Starting with proportion of political

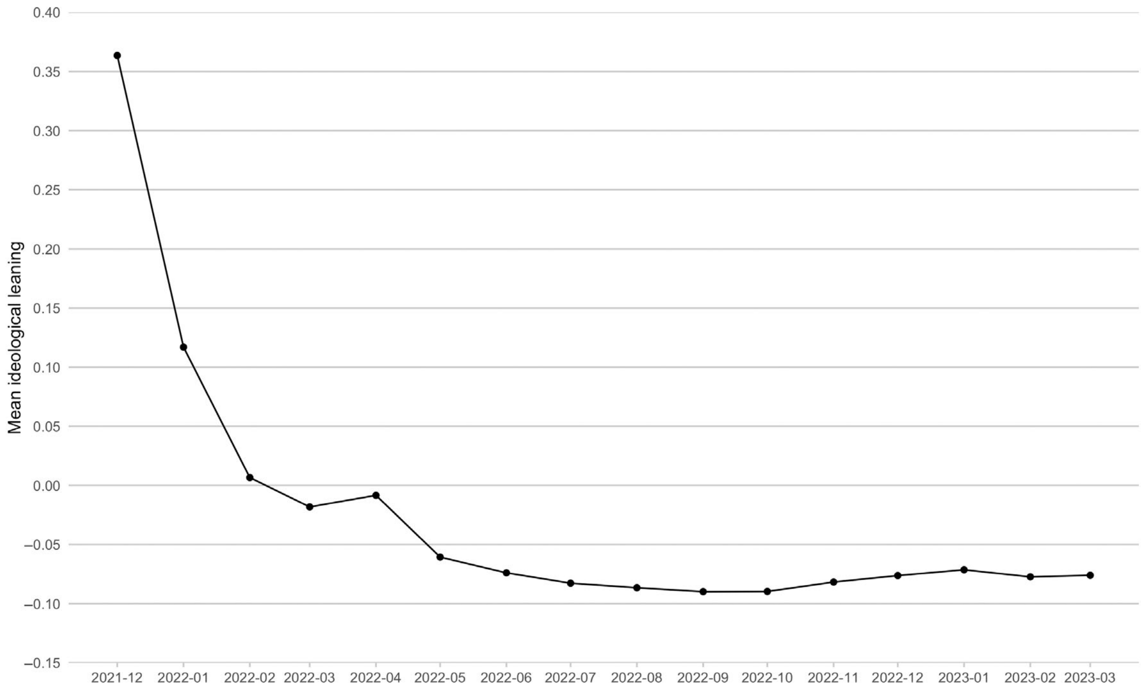

FIGURE 6 Average monthly ideology scores.

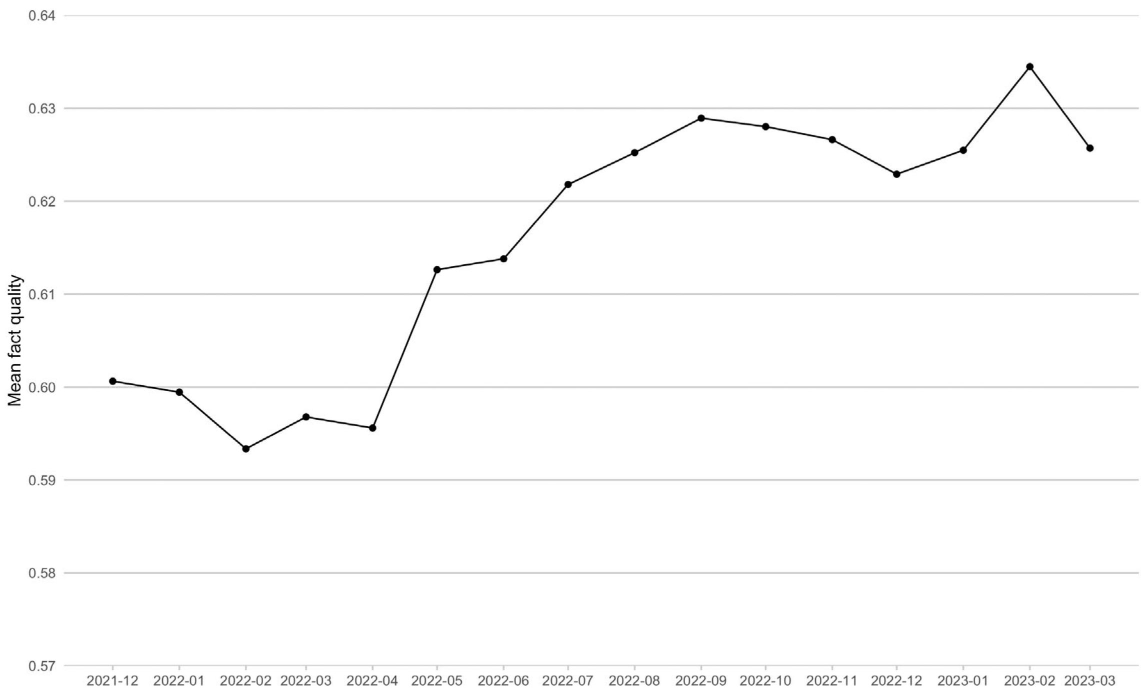

FIGURE 7 Average monthly fact quality scores.

(MBFC-classified) content, panel 8a shows that the more participants reported using "mainstream media" and the older they were, the more political content their PIEs contained. Moving to ideological leaning (panel 8b), we see that the more satisfied participants reported being with mainstream media, the less conservative content their feeds contained. In contrast, mainstream media use frequency, conservative self-identification, strength of Republican Party identification, age, and belief that the January 6th insurrection was exaggerated all predicted more conservative content. Finally, panel 8c reveals that frequency of mainstream media consumption, frequency of X/Twitter use, and belief that January 6 was exaggerated were all associated with lower mean fact quality

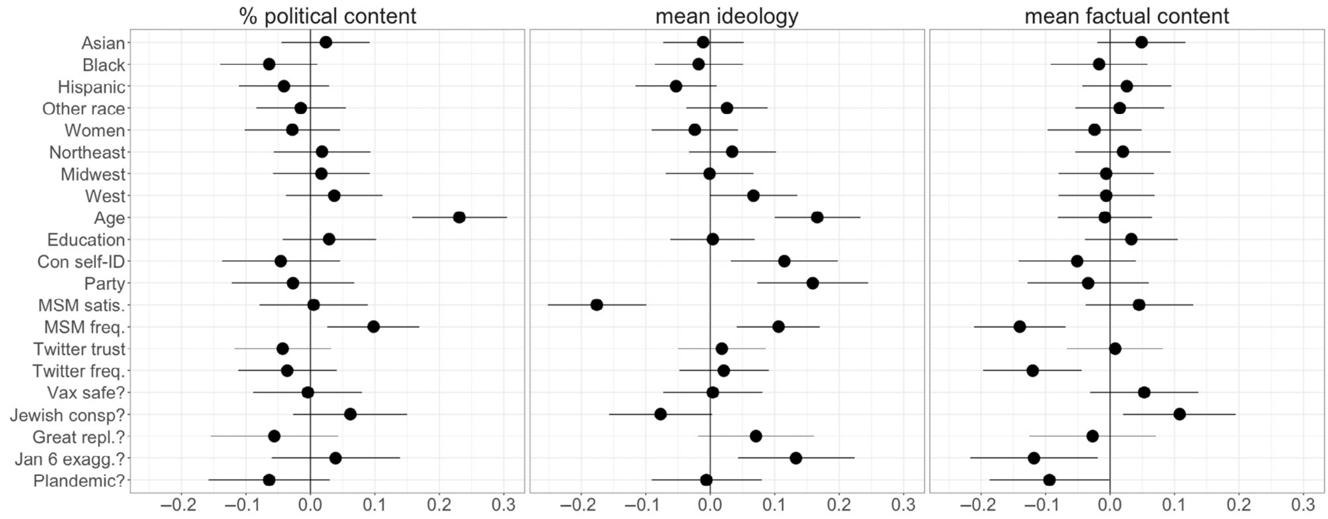

FIGURE 8 Regression models predicting proportion of MBFC-coded ("political") posts, mean ideology, and mean fact quality. Regression coefficients are standardized. See SI section C in Data S1 for the complete phrasing of these questions. SI section E in Data S1 presents this data in tabular format.

scores. However, unexpectedly, the belief that Jewish people control the US government and media system predicts higher mean fact quality.

8 | DISCUSSION

Our analyses yield some results consistent with previous research along with a few surprises. The total number of unique posts we found lacks comparable figures in the literature—we cannot know whether it is a relatively large or small amount compared either to X/Twitter in the past or to other platforms. We can say, however, that previous user-centric studies have found that around 2%–3% of content that participants view is political (Cronin et al., 2022; Wojcieszak et al., 2022). The fact that over 10% of our links were classified as politically relevant by MBFC indicates that people may be exposed to more political content briefly in their feeds than direct website visits would indicate. We also found that our sample's ideological mean and distribution lean to the left, which is consistent with what we know about Twitter's user base (Freelon, 2019; Wojcik & Hughes, 2019), and that low-quality content is fairly rare (Allcott & Gentzkow, 2017; Grinberg et al., 2019; Guess et al., 2019).

While we did not foresee Elon Musk's purchase of X/Twitter occurring during our study period, it offered an opportunity to observe how our participants' environments changed pre- and post-acquisition. Post volume across all timelines began decreasing before the purchase and continued after, suggesting that Musk probably was not the sole cause of the decrease, although his corporate decisions and rhetoric may have contributed to it. A longer study period would have helped us determine how anomalous the dip in volume starting in August 2022 was. Interestingly, none of the three main outcome variables (political content percentage, mean ideological leaning, and mean fact quality) manifested any recognizable difference after October 2022. Generally, each quantity attained stability after all participants had been recruited at the end of May. It seems Musk's purchase did not result in a marked shift to right-wing or low-quality political hyperlinks (although this may result from the leftward skew of our sample). This is somewhat inconsistent with non-peer-reviewed analyses showing post-acquisition increases in hate speech (Center for Countering Digital Hate, 2022; Siddiqui & Merrill, 2023) and misinformation (Carniel, 2023). More comprehensive research must be conducted to answer these questions conclusively. Broadly, our results highlight the importance of extended study periods to document the influence of major changes in platform policy and leadership (see, e.g., Cavusoglu et al., 2016).

Our statistical models support previous findings linking age, media use frequency, and belief in certain conspiracy theories (but not others) to our outcome variables. Older individuals have long been known to consume more political and right-wing content than their younger counterparts (Guess et al., 2019, 2021; Osmundsen et al., 2021). However, contrary to these studies, age here does not predict the presence of low-quality links in information environments. Those who state that they consume mainstream media content more frequently have more political content, more right-wing content, and lower-quality content in their PIEs than those who consume less mainstream content. This is consistent with research showing that people with higher levels of political interest and knowledge may be more likely to consume and spread misinformation if it helps them defend their opinions (Valenzuela et al., 2019). Further, our findings suggest that not all conspiracy theories are connected equally strongly to right-wing or low-quality sites. Of the four conspiracy theories we included as predictors, only belief in the notion that the mainstream media exaggerated the January 6 insurrection to disparage Donald Trump predicted more right-wing PIE content. This belief also predicted lower fact quality, but the antisemitic belief that Jewish people control American media and politics actually predicted higher fact quality. This may be evidence of a monitorial impulse among conspiracy and misinformation consumers—although they dislike and disbelieve what we understand as high-quality media outlets, they may consume their output nonetheless to decode and denounce what they perceive as lies (cf. Munyaka et al., 2022) or cite them misleadingly to support their baseless claims (Aaronovitch, 2010).

Overall, this study offers a thoroughly moderate and apolitical impression of our participants' X/Twitter information environments. News and politics constitute a minor concern, with 90% of PIE content devoted to other topics (which we plan on exploring in future research). Ideologically, most timelines stick fairly close to the center, with very few occupying what we might consider extreme echo chambers (see Bruns, 2019; Guess et al., 2018). In terms of fact quality, the aggregate mean hews close to the "mostly factual" rating, and extremely low- and high-quality content are both rare. And all these trends hold steady for the better part of a year. Our statistical models reveal no major asymmetries for any of the demographic variables we tested. Normatively, this could be cause for optimism, as it seems few have opted into substantial amounts of politically extreme or chronically low-quality content. The statistical relationships we found between ideological and party self-identification on the survey and mean ideological leaning of timeline content suggests that our participants' X/Twitter following patterns index their offline political identities fairly closely. In other words, it is reasonable to assume that the content people choose to view on X/Twitter probably resembles what they choose in other media contexts.

Yet our methodological reliance on hyperlinks means that the empirical impression we offer here is incomplete. While lists of unreliable domains and MBFC-style domain-based ideology and fact quality ratings have been used in many studies (Grinberg et al., 2019; Guess et al., 2019, 2021), it is possible that important political content is hiding somewhere beyond these methods' ability to capture. They account neither for political discussions that lack hyperlinks, for example, nor for visual or audio content that may be political in nature. A more comprehensive search for political content would be a logical next step for PIE research, but the necessary methods will be much more labor-intensive than cross-referencing a pre-classified list of web domains. Even if we were to focus exclusively on text, we might need a lengthy dictionary of political terms or a supervised machine learning model trained on posts classified as "political." While hyperlink analysis offers a starting point, future research should broaden its search parameters for political content.

Our study has several other limitations that should be acknowledged. First, it is limited to X/Twitter, which does not represent anyone's complete PIE. Nevertheless, the ideological preferences that emerged in our participants' X/Twitter timelines were strongly correlated with their survey answers, suggesting that they are not limited to X/Twitter. Second, we were not able to determine what our participants viewed or clicked on, unlike Comscore and Netquest data which show clicks but not the broader hyperlink universe from which they were drawn. It would be ideal to integrate both types of data in a single study, but we are not aware of any tools capable of doing so. Third, while many of the potential participants we solicited were willing to fill out our survey, only a small proportion thereof were willing to grant access to their X/Twitter feeds. While we did not investigate what factors might have driven potential participants' decisions to participate or not, existing research indicates that privacy concerns, perceived relevance of the research, and incentive size may be relevant (Pfiffner & Friemel, 2023; Silber et al., 2022). We further note that while our final sample matches the US Census fairly closely for the most part (see discussion in SI section B in Data S1), our findings will not necessarily generalize to populations outside the United States.

Although PIEGraph offers one level of understanding of our participants' respective X/Twitter feeds, it does not address the question of how X/Twitter's algorithms impact said content. Specifically, because the API we used delivers all content from the set of users each participant follows, individual choices to use X/Twitter's "For You" page vs. its "Following" page would not affect our findings. Whether this should be considered a limitation is debatable—obviously it would be worthwhile to be able to distinguish algorithmically-delivered content from what users chose themselves, but X/Twitter has never offered this data to outside researchers. While X/Twitter's algorithm curation is unlikely to affect the sets of users our participants follow (the present object of study), it probably influences what they see while using the platform. Earlier studies have shown that right-leaning US media outlets were amplified more than left-leaning ones by X/Twitter's algorithm (Huszár et al., 2022), and that posts containing external links were considerably less prevalent in the algorithmic timelines than in the chronological ones (Bandy & Diakopoulos, 2021). Before the recent changes to X/Twitter's API, algorithmic audit approaches based on synthetic or "sock puppet" accounts were used to collect algorithmic content (Bandy & Diakopoulos, 2021; Bartley et al., 2021). However, X/Twitter's API policies severely limited the number of accounts from which they could collect data—these two studies each employed only eight accounts. Ideally, social media platforms would include algorithmic metadata in their API output to help determine the effects of algorithms on the content of an individual's feed, but no API of which we are currently aware does so.

More generally, user-centric approaches, while useful for answering certain research questions, are not a methodological panacea. The process of interpreting digital traces as indicators of human intention and attention requires many assumptions, some of which will not hold true in all circumstances. For example, any system that tracks PIEs or interactive online behavior (clicks, shares, etc.) can tell us what a user did or might have seen, but they cannot definitively tell us why. We can say, in other words, that a particular piece of content appeared in a user's environment or that they clicked on it (depending on the system), but we cannot definitively know that they viewed, read, agreed with, or interpreted it as we would have. Moreover, systems like ours rely on data provided by platform companies, which is incomplete in the best-case scenario and to which access can be revoked at any time (see discussion below). Other systems that do not rely on such data must use only what can be extracted from the webpage's source code or screenshots, which may exclude valuable data such as hyperlinks and comments that load automatically as the user scrolls down. Despite these limitations, we maintain that user-centric approaches in general, and ours in particular, can yield useful insights about the content streams people encounter through social media.

Our overall approach and findings are relevant to multiple of the broader aspects of the information ecosystem, particularly government policy and the technological design choices of social media platforms. The First Amendment to the US Constitution broadly limits the power of American law to compel social media companies to operate their platforms in any particular manner (Hooker, 2019). Nevertheless, laws passed in Texas and Florida in 2021 attempt to prevent platforms from banning or otherwise punishing accounts based on their political viewpoints. The Supreme Court has decided to rule on the constitutionality of such measures, but research like ours can help reveal the extent to which (1) platforms' terms-of-service enforcement and (2) so-called anti-censorship laws affect the visibility of affected accounts. In particular, our methods could have tracked the prevalence of accounts known or suspected of spreading misinformation or hateful content over time and on a state-by-state basis. Just as the current study notes a marked decline in content volume after Musk's takeover of X/Twitter, it could have also documented content changes following the implementation of state and federal laws as well as new platform policies, had the company not decided to place data access beyond the reach of most researchers.

Between January 2021 and April 2023, X/Twitter offered academic researchers free access to 10 million posts per month with the submission of a brief application. Elon Musk's decision to end this program and charge $100 per month for the lowest tier of data access, which as of this writing offers only 10,000 posts per month, drastically limits both the kinds of studies that can be rigorously conducted on X/Twitter as well as the population of researchers that can conduct them. To put it plainly, while we paid nothing to X/Twitter to produce the research described in this article, PIEGraph is no longer usable without paying at least $5000 per month, and possibly more. Other platforms have made different decisions with respect to data availability: while Reddit has sharply reduced the utility of its data by requiring short-lived API tokens (Pushshift-Support, 2023), Meta has opened access to data about Facebook's public groups and pages via a partnership with the University of Michigan (Meta, 2023). Social media data access in the United States is currently subject to the whims of company leadership, rendering the study of content quality—along with every other social media topic—indefinitely tenuous. This, in turn, places a critical aspect of information ecosystem health at existential risk.

The importance of research on digital harms such as misinformation has been recognized by the European Union, which directs "very large online platforms" to provide data to independent researchers "for the sole purpose of conducting research that contributes to the detection, identification, and understanding of systemic risks in the Union" (Digital Services Act, 2022). We believe a strong case can be made that the research presented here addresses a specific "systemic risk" as the EU defines it: low-quality information. Therefore, in addition to this study's inherent benefits, it stands as an example of the acknowledged empirical value of social media research that currently lies beyond reach (at least for X/Twitter) under the US's unregulated data access status quo. That value could be reclaimed with regulation similar to the EU's Digital Services Act, which imposes on platforms a very different set of public-minded requirements than the Florida and Texas laws discussed earlier. A crucial step in moving toward solutions for digital risks, and a healthier information ecosystem more broadly, is mapping risk prevalence, and we believe clear government policy is the best way to ensure access to social media data (both platform- and user-centric) consistently across platforms.

FUNDING INFORMATION

This research was funded by the John S. and James L. Knight Foundation and the William and Flora Hewlett Foundation.

ENDNOTES

[1] MBFC's ideology scale has 67 possible ordinal categories: −33 through −1 for left of center and 1 through 33 for right of center, with 0 lying exactly in the middle. We normalized these to range between −1 and 1. Its factual reporting scale has six possible ordinal categories—Very High, High, Mostly Factual, Mixed, Low, and Very Low—and we normalized these to range between 0 and 1.

[2] In this count, each unique post ID counts once, where uniqueness is considered across all PIEs. In other words, a single post appearing in 10 different participant timelines would count only once.

[3] In these counts, uniqueness applies within PIEs, such that a single post ID would count once for each timeline in which it appeared.

REFERENCES

Aaronovitch, D. (2010). *Voodoo histories: The role of the conspiracy theory in shaping modern history* (1st ed.). Riverhead Hardcover.

Aires, V. P., Nakamura, G., & Nakamura, E. F. (2019). A link-based approach to detect media bias in news websites. In Companion proceedings of the 2019 world wide web conference, pp. 742–745. https://doi.org/10.1145/3308560.3316460

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, *6*(14), eaay3539. https://doi.org/10.1126/sciadv.aay3539

Altay, S., Nielsen, R. K., & Fletcher, R. (2022). Quantifying the "infodemic": People turned to trustworthy news outlets during the 2020 coronavirus pandemic. *Journal of Quantitative Description: Digital Media*, *2*. https://doi.org/10.51685/jqd.2022.020

Araujo, T., Ausloos, J., van Atteveldt, W., Loecherbach, F., Moeller, J., Ohme, J., Trilling, D., van de Velde, B., DE Vreese, C., & Welbers, K. (2022). OSD2F: An open-source data donation framework. *Computational Communication Research*, *4*(2), 372–387. https://doi.org/10.5117/CCR2022.2.001.ARAU

Bandy, J., & Diakopoulos, N. (2021). More accounts, fewer links: How algorithmic curation impacts media exposure in twitter timelines. In Proceedings of the ACM on human-computer interaction, 5(CSCW1), pp. 78:1–78:28. https://doi.org/10.1145/3449152

Bartley, N., Abeliuk, A., Ferrara, E., & Lerman, K. (2021). Auditing algorithmic bias on twitter. In 13th ACM web science conference 2021, pp. 65–73. https://doi.org/10.1145/3447535.3462491

Bayer, J. B., Triệu, P., & Ellison, N. B. (2020). Social media elements, ecologies, and effects. *Annual Review of Psychology*, *71*(1), 471–497. https://doi.org/10.1146/annurev-psych-010419-050944

Beckers, K., Van Aelst, P., Verhoest, P., & d'Haenens, L. (2021). What do people learn from following the news? A diary study on the influence of media use on knowledge of current news stories. *European Journal of Communication*, *36*(3), 254–269.

Bentley, F., Quehl, K., Wirfs-Brock, J., & Bica, M. (2019). Understanding online news behaviors. In Proceedings of the 2019 CHI conference on human factors in computing systems, pp. 1–11. https://doi.org/10.1145/3290605.3300820

Breuer, J., Kmetty, Z., Haim, M., & Stier, S. (2022). User-centric approaches for collecting Facebook data in the 'post-API age': Experiences from two studies and recommendations for future research. *Information, Communication & Society*, 26(14), 2649–2668. https://doi.org/10.1080/1369118X.2022.2097015

Bruns, A. (2019). *Are filter bubbles real?* John Wiley & Sons.

Calvillo, D. P., Ross, B. J., Garcia, R. J. B., Smelter, T. J., & Rutchick, A. M. (2020). Political ideology predicts perceptions of the threat of COVID-19 (and susceptibility to fake news about it). *Social Psychological and Personality Science*, *11*, 1119–1128. https://doi.org/10.1177/1948550620940539

Carniel, B. (2023). Misinformation superspreaders are thriving on Musk-owned Twitter. Health Feedback. https://healthfeedback.org/misinformation-superspreaders-thriving-on-musk-owned-twitter/

Carter, M. C., Cingel, D. P., Ruiz, J. B., & Wartella, E. (2023). Social media use in the context of the personal social media ecosystem framework. *Journal of Communication*, *73*(1), 25–37. https://doi.org/10.1093/joc/jqac038

Carter, S., & Mankoff, J. (2005). When participants do the capturing: The role of media in diary studies. In Proceedings of the SIGCHI conference on human factors in computing systems, pp. 899–908. https://doi.org/10.1145/1054972.1055098

Cavusoglu, H., Phan, T. Q., Cavusoglu, H., & Airoldi, E. M. (2016). Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. *Information Systems Research*, *27*(4), 848–879. https://doi.org/10.1287/isre.2016.0672

Center for Countering Digital Hate. (2022). The Musk Bump: Quantifying the rise in hate speech under Elon Musk. https://counterhate.com/blog/the-musk-bump-quantifying-the-rise-in-hate-speech-under-elon-musk/

Chen, M., Chu, X., & Subbalakshmi, K. P. (2022). MMCoVaR: Multimodal COVID-19 vaccine focused data repository for fake news detection and a baseline architecture for classification. In Proceedings of the 2021 IEEE/ACM international conference on advances in social networks analysis and mining, pp. 31–38. https://doi.org/10.1145/3487351.3488346

Christner, C., Urman, A., Adam, S., & Maier, M. (2022). Automated tracking approaches for studying online media use: A critical review and recommendations. *Communication Methods and Measures*, *16*(2), 79–95.

Cronin, J., von Hohenberg, B. C., Gonçalves, J. F. F., Menchen-Trevino, E., & Wojcieszak, M. (2022). The (null) over-time effects of exposure to local news websites: Evidence from trace data. *Journal of Information Technology & Politics*, *20*(4), 407–421. https://doi.org/10.1080/19331681.2022.2123878

Digital Services Act. (2022). § Article 40, Section 4. https://www.eu-digital-services-act.com/Digital_Services_Act_Article_40.html

Farhall, K., Carson, A., Wright, S., Gibbons, A., & Lukamto, W. (2019). Political elites' use of fake news discourse across communications platforms. *International Journal of Communication*, *13,* 4353–4375.

Fedeli, S., & Matsa, K. E. (2018). *Use of mobile devices for news continues to grow, outpacing desktops and laptops*. Pew Research Center.

Freelon, D. (2019). *Tweeting left, right, & center: How users and attention are distributed across twitter* (pp. 1–38). John S. & James L. Knight Foundation.

Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, *37*(2), 145–156. https://doi.org/10.1080/10584609.2020.1723755

Geer, J. G., & Kahn, K. F. (1993). Grabbing attention: An experimental investigation of headlines during campaigns. *Political Communication*, *10*(2), 175–191.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guess, A. M. (2021). (Almost) everything in moderation: New evidence on Americans' online media diets. *American Journal of Political Science*, *65*(4), 1007–1022. https://doi.org/10.1111/ajps.12589

Guess, A. M., Aslett, K., Tucker, J., Bonneau, R., & Nagler, J. (2021). Cracking open the news feed: Exploring what U.S. Facebook users see and share with large-scale platform data. *Journal of Quantitative Description: Digital Media*, *1*. https://doi.org/10.51685/jqd.2021.006

Guess, A. M., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), eaau4586. https://doi.org/10.1126/sciadv.aau4586

Guess, A. M., Nyhan, B., Lyons, B., & Reifler, J. (2018). Avoiding the echo chamber about echo chambers. *Knight Foundation*, *2*, 1–25.

Gulyas, A., O'Hara, S., & Eilenberg, J. (2019). Experiencing local news online: Audience practices and perceptions. *Journalism Studies*, *20*(13), 1846–1863. https://doi.org/10.1080/1461670X.2018.1539345

Haim, M., & Nienierza, A. (2019). Computational observation: Challenges and opportunities of automated observation within algorithmically curated media environments using a browser plug-in. *Computational Communication Research*, *1*(1), 79–102. https://doi.org/10.5117/CCR2019.1.004.HAIM

Halavais, A. (2019). Overcoming terms of service: A proposal for ethical distributed research. *Information, Communication & Society*, *22*(11), 1567–1581. https://doi.org/10.1080/1369118X.2019.1627386

Hooker, M. P. (2019). Censorship, Free Speech & Facebook: Applying the first amendment to social media platforms via the public function exception. *Washington Journal of Law, Technology & Arts*, *15*(1), 36–73.

Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on twitter. *Proceedings of the National Academy of Sciences*, *119*(1), e2025334119. https://doi.org/10.1073/pnas.2025334119

Janét, K., Richards, O., & Landrum, A. R. (2022). Headline format influences evaluation of, but not engagement with, environmental news. *Journalism Practice*, *16*(1), 35–55. https://doi.org/10.1080/17512786.2020.1805794

Knight Foundation. (2022). Media and democracy: Unpacking America's complex views on the digital Public Square. https://knightfoundation.org/reports/media-and-democracy/

Kucharski, A. (2016). Study epidemiology of fake news. *Nature*, *540*(7634), 7634. https://doi.org/10.1038/540525a

Lin, H., Lasser, J., Lewandowsky, S., Cole, R., Gully, A., Rand, D., & Pennycook, G. (2022). High level of agreement across different news domain quality ratings. *PsyArXiv*. https://doi.org/10.31234/osf.io/qy94s

Menchen-Trevino, E. (2016). Web historian: Enabling multi-method and independent research with real-world web browsing history data. In iConference 2016 proceedings. http://hdl.handle.net/2142/89388

Meta. (2023). Meta content library and API | transparency center. https://transparency.fb.com/researchtools/meta-content-library/

Mihelj, S., Kondor, K., & Štětka, V. (2022). Audience engagement with COVID-19 news: The impact of lockdown and live coverage, and the role of polarization. *Journalism Studies*, *23*(5–6), 569–587. https://doi.org/10.1080/1461670X.2021.1931410

Moe, H., & Ytre-Arne, B. (2022). The democratic significance of everyday news use: Using diaries to understand public connection over time and beyond journalism. *Digital Journalism*, *10*(1), 43–61. https://doi.org/10.1080/21670811.2020.1850308

Munyaka, I., Hargittai, E., & Redmiles, E. (2022). The misinformation paradox: Older adults are cynical about news media, but engage with it anyway. *Journal of Online Trust and Safety*, *1*(4), 4. https://doi.org/10.54501/jots.v1i4.62

Narayan, B., Case, D. O., & Edwards, S. L. (2011). The role of information avoidance in everyday-life information behaviors. *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1–9. https://doi.org/10.1002/meet.2011.14504801085

Nelson, J. L., & Lei, R. F. (2018). The effect of digital platforms on news audience behavior. *Digital Journalism*, *6*(5), 619–633. https://doi.org/10.1080/21670811.2017.1394202

Nelson, J. L., & Lewis, S. C. (2022). The structures that shape news consumption: Evidence from the early period of the COVID-19 pandemic. *Journalism*, *23*, 2495–2512. https://doi.org/10.1177/14648849221095335

Nelson, J. L., & Taneja, H. (2018). The small, disloyal fake news audience: The role of audience availability in fake news consumption. *New Media & Society*, *20*, 3720–3737. https://doi.org/10.1177/1461444818758715

Nelson, J. L., & Webster, J. G. (2017). The myth of partisan selective exposure: A portrait of the online political news audience. *Social Media + Society*, *3*(3). https://doi.org/10.1177/2056305117729314

Ng, L. H. X., & Loke, J. Y. (2021). Analyzing public opinion and misinformation in a COVID-19 telegram group chat. *IEEE Internet Computing*, *25*(2), 84–91. https://doi.org/10.1109/MIC.2020.3040516

Norris, T. B., & Suomela, T. (2017). Information in the ecosystem: Against the "information ecosystem". *First Monday, 22*(9). https://doi.org/10.5210/fm.v22i9.6847

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2023). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Communication Methods and Measures*, 1–18. https://doi.org/10.1080/19312458.2023.2181319

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *American Political Science Review*, *115*(3), 999–1015. https://doi.org/10.1017/S0003055421000290

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin.

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 11. https://doi.org/10.1038/s41562-021-01117-5

Pfiffner, N., & Friemel, T. N. (2023). Leveraging data donations for communication research: Exploring drivers behind the willingness to donate. *Communication Methods and Measures*, *17*(3), 227–249. https://doi.org/10.1080/19312458.2023.2176474

Pushshift-Support. (2023). Pushshift live again and how moderators can request Pushshift access [Reddit post]. R/Pushshift. www.reddit.com/r/pushshift/comments/14ei799/pushshift_live_again_and_how_moderators_can/

Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., Chiatti, A., Cho, M., Roehrick, K., Yang, X., Gagneja, A., Brinberg, M., Muise, D., Lu, Y., Luo, M., Fitzgerald, A., & Yeykelis, L. (2021). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. *Human–Computer Interaction*, *36*(2), 150–201. https://doi.org/10.1080/07370024.2019.1578652

Ribeiro, F., Henrique, L., Benevenuto, F., Chakraborty, A., Kulshrestha, J., Babaei, M., & Gummadi, K. (2018). Media bias monitor: Quantifying biases of social media news outlets at large-scale. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1), 290–299.

Rieh, S. Y., Kim, Y.-M., Yang, J. Y., & St Jean, B. (2010). A diary study of credibility assessment in everyday life information activities on the web: Preliminary findings. *Proceedings of the American Society for Information Science and Technology*, *47*(1), 1–10. https://doi.org/10.1002/meet.14504701182

Saltz, E., Leibowicz, C. R., & Wardle, C. (2021). Encounters with visual misinformation and labels across platforms: An interview and diary study to inform ecosystem approaches to misinformation interventions. In Extended abstracts of the 2021 CHI conference on human factors in computing systems, pp. 1–6. https://doi.org/10.1145/3411763.3451807

Siddiqui, F., & Merrill, J. B. (2023). Elon Musk's twitter pushes hate speech, extremist content into 'for you' pages. Washington Post. https://www.washingtonpost.com/technology/2023/03/30/elon-musk-twitter-hate-speech/

Silber, H., Breuer, J., Beuthner, C., Gummer, T., Keusch, F., Siegers, P., Stier, S., & Weiß, B. (2022). Linking surveys and digital trace data: Insights from two studies on determinants of data sharing behaviour. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(suppl 2), S387–S407. https://doi.org/10.1111/rssa.12954

Stier, S., Mangold, F., Scharkow, M., & Breuer, J. (2022). Post post-broadcast democracy? News exposure in the age of online intermediaries. *American Political Science Review*, *116*(2), 768–774. https://doi.org/10.1017/S0003055421001222

Sunstein, C. (2018). *#Republic: Divided democracy in the age of social media*. Princeton University Press.

Thorson, K., & Wells, C. (2016). Curated flows: A framework for mapping media exposure in the digital age. *Communication Theory*, *26*(3), 309–328. https://doi.org/10.1111/comt.12087

Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism*, *7*(6), 802–823. https://doi.org/10.1080/21670811.2019.1623701

Vandewater, E. A., & Lee, S.-J. (2009). Measuring children's media use in the digital age: Issues and challenges. *American Behavioral Scientist*, *52*, 1152–1176. https://doi.org/10.1177/0002764209331539

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, *37*(3), 350–375.

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking* (Vol. 27). Council of Europe Strasbourg.

Weeks, B. E., Menchen-Trevino, E., Calabrese, C., Casas, A., & Wojcieszak, M. (2021). Partisan media, untrustworthy news sites, and political misperceptions. *New Media & Society*, *25*(10), 2644–2662. https://doi.org/10.1177/14614448211033300

Williamson, P. (2016). Take the time and effort to correct misinformation. *Nature*, *540*(7632), 171. https://doi.org/10.1038/540171a

Wojcieszak, M., Clemm von Hohenberg, B., Casas, A., Menchen-Trevino, E., de Leeuw, S., Gonçalves, A., & Boon, M. (2022). Null effects of news exposure: A test of the (un)desirable effects of a 'news vacation' and 'news binging'. *Humanities and Social Sciences Communications*, *9*(1), 1–10. https://doi.org/10.1057/s41599-022-01423-x

Wojcik, S., & Hughes, A. (2019). How twitter users compare to the general public. Pew Research Center. https://www.pewinternet.org/2019/04/24/sizing-up-twitter-users/

Yang, T., Majó-Vázquez, S., Nielsen, R. K., & González-Bailón, S. (2020). Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences*, *117*(46), 28678–28683. https://doi.org/10.1073/pnas.2006089117

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Freelon, D., Pruden, M. L., Malmer, D., Wu, Q., Xia, Y., Johnson, D., Chen, E., & Crist, A. (2024). What's in your PIE? Understanding the contents of personalized information environments with PIEGraph. *Journal of the Association for Information Science and Technology*, 1–15. https://doi.org/10.1002/asi.24869