



Article

# The puzzle of misinformation: Exposure to unreliable content in the United States is higher among the better informed

new media & society

1–18

© The Author(s) 2023



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/14614448231196863

[journals.sagepub.com/home/nms](https://journals.sagepub.com/home/nms)



**Alvin Zhou** 

University of Minnesota, USA

**Tian Yang** 

The Chinese University of Hong Kong, Hong Kong SAR

**Sandra González-Bailón**

University of Pennsylvania, USA

## Abstract

Healthy news consumption requires limited exposure to unreliable content and ideological diversity in the sources consumed. There are two challenges to this normative expectation: the prevalence of unreliable content online; and the prominence of misinformation within individual news diets. Here, we assess these challenges using an observational panel tracking the browsing behavior of  $N \approx 140,000$  individuals in the United States for 12 months (January–December 2018). Our results show that panelists who are exposed to misinformation consume more reliable news and from a more ideologically diverse range of sources. In other words, exposure to unreliable content is higher among the better informed. This association persists after we control for partisan leaning and consider inter- and intra-person variation. These findings highlight the tension between the positive and negative consequences of increased exposure to news content online.

---

## Corresponding author:

Sandra González-Bailón, Annenberg School for Communication, University of Pennsylvania, 3620 Walnut Street, Philadelphia, PA 19104, USA.

Email: [sgonzalezbailon@asc.upenn.edu](mailto:sgonzalezbailon@asc.upenn.edu)

## Keywords

Political Information, media environment, misinformation, news exposure, news use, web data

The circulation of misinformation through online networks is one of the most salient challenges confronting researchers and policymakers today. The current media environment allows the unmonitored publication of content that online networks help diffuse and amplify. The visibility of information online results from a combination of social and algorithmic processes that often introduce biases in the content people see (Huszár et al., 2022). Curation mechanisms act as mediators channeling traffic to sources that would have a smaller audience otherwise (Nielsen and Fletcher, 2022; Scharkow et al., 2020). On the one hand, misinformation benefits from the amplification effects of social and algorithmic curation, and from the use of clickbait and emotional triggers (Benkler et al., 2018; Chen et al., 2015; Vosoughi et al., 2018). On the other hand, only a minority of people actively circulate and consume misinformation (at least, as assessed using data from the United States, e.g. Allen et al., 2020; Grinberg et al., 2019; Guess et al., 2020b; Nelson and Taneja, 2018). Most of the information people consume online still comes from reliable news sources (Allen et al., 2020; Altay et al., 2022; Yang et al., 2020). Misinformation, in other words, seems contained within small clusters of specific groups.

The question of reach, however, is separate from the question of impact. If, for instance, the pathways that lead to misinformation reveal partisan divides (i.e. if Republicans are more likely to engage with unreliable content, as prior research shows, Allcott and Gentzkow, 2017; Guess et al., 2019), then the small group of people consuming misinformation may still have a disproportionate impact on the politics of attention and polarizing group dynamics. Likewise, if age is the main demographic factor underlying engagement with unreliable content (e.g. Guess et al., 2019), and older groups also exhibit more intense forms of political involvement (e.g. Krupnikov and Ryan, 2022), misinformation may again have an influence disproportional to its reach. Thus, the main issue is not just how prevalent misinformation is but how it fits into people's broader news diets. If unreliable content dominates news diets, its impact is likely to be higher than if it is a small part of a broader, more reliable, and ideologically diverse set of news.

Existing studies indeed show that people consuming unreliable content also consume a large amount of reliable news (e.g. Guess et al., 2021; Nelson and Taneja, 2018). Past research also shows that exposure to misinformation is fairly concentrated, with a small number of people driving most of that exposure (e.g. Allen et al., 2020; Grinberg et al., 2019; Watts et al., 2021). Here we provide additional analyses that support these past findings while casting novel light on the relationship between the diversity of news diets and exposure to unreliable content.

The literature on selective exposure, along with research on the hyper-partisan nature of misinformation (e.g. Eady et al., 2023; Guay et al., 2022), suggest that those consuming unreliable content seek ideologically congruent information and should therefore have less ideologically diverse news diets. On the other hand, research also suggests that misinformation consumers exhibit higher levels of political interest (e.g. Nelson and

Taneja, 2018; Pennycook and Rand, 2021): they are more voracious in their news consumption and visit a wider range of information sources, potentially increasing the ideological diversity of their news inventories.

Our analyses aim to identify which of these alternative scenarios receives empirical support: is exposure to unreliable content associated with more ideologically congruent diets? Or is it associated with more diverse news consumption? To address these questions, we use uniquely rich panel data tracking the web browsing behavior of more than 100,000 panelists based in the United States for a period of 12 months (January–December 2018). The data contain the URLs visited by the panelists as well as their demographic attributes, which allows us to identify behavioral and demographic profiles associated with exposure to misinformation.

Before introducing our data, the following section offers a more detailed discussion of prior research and what it tells us about the prevalence of unreliable content and the audiences more likely to engage with it. We pay special attention to observational studies that analyze the relative volume of misinformation compared with reliable news, and studies that focus on the individual-level correlates of observed exposure to that content. We then describe our panel and how it alleviates some of the limitations of prior work, especially when it comes to answering our main guiding question: what other news do people engaging with misinformation consume? In answering this question, we discuss the puzzle of misinformation. The consumption of unreliable content is just one piece in the larger picture of broader news consumption and, when analyzed, this larger picture reveals an apparent contradiction: that exposure to unreliable content is higher among the better informed. In the context of our data, “better informed” is defined as having more exposure to reliable news and a more ideologically diverse news diet.

The puzzle we uncover is similar in nature to the tension between positive and negative engagement that has been studied on social media, i.e. the fact that beneficial behaviors, like increased participation in political talk, are also associated with dysfunctional consequences, like sharing more misinformation and relying on partisan outlets (Fletcher et al., 2023; Rossini et al., 2021; Valenzuela et al., 2019). It is also similar to the observed contradiction between the cynical attitudes toward news media that older adults declare and their tendency to still read and share news they distrust (Munyaka et al., 2022). Unpacking this puzzle, we argue, is important to inform interventions designed to curtail the effects of misinformation. For instance, approaches focused on increasing literacy or promoting cross-cutting, diverse exposure will not produce the theorized effects if a higher engagement with diverse, reliable news is positively associated with exposure to unreliable content. We conclude by advocating for a systemic approach to addressing the problem of misinformation (as opposed to individual approaches centered on psychological mechanisms). We also revisit the concept of informed citizenry, and what it means in the context of the current news ecosystem.

## The study of misinformation

The analysis of misinformation and so-called “fake news” has been of special interest to academics and journalists since 2016. That year, two high-profile political events—the US Presidential Election and the Brexit referendum—made patent the risks that online

networks create in allowing the unmonitored dissemination of false and inaccurate content (Lazer et al., 2018). These risks were also seen and denounced in many other political contexts around the world, especially after the COVID-19 pandemic created new fronts on which to fight misinformation (Zarocostas, 2020). A common fear among observers and analysts is the impact that misinformation may have on election results, compliance with policy recommendations and, more generally, the legitimacy and stability of the democratic process. Democracies need informed citizens to ensure that accountability mechanisms work. Misinformation not only weakens this foundation; it also threatens to aggravate conflict through misrepresentations and inaccurate portrayals of political and social realities.

The number of empirical articles that have been published on the topic of misinformation has grown substantially in the past 6 years, but the reported results fall in three broad categories: descriptive statistics measuring the prevalence and reach of misinformation (i.e. Allen et al., 2020; Altay et al., 2022; Grinberg et al., 2019; Guess et al., 2020b; Nelson and Taneja, 2018; Vosoughi et al., 2018; Watts et al., 2021); demographic characteristics of misinformation's likely audiences (e.g. Allcott and Gentzkow, 2017; Guess et al., 2019, 2021); and the impact of interventions designed to curtail or correct misinformation (e.g. Aslett et al., 2022; Pennycook and Rand, 2019; Vraga et al., 2022; Walter et al., 2020). The studies differ in their empirical focus (i.e. Twitter, Facebook, YouTube, the web) and their measures of engagement with misinformation (e.g. they track exposure to web domains or engagement in the form of sharing on social media platforms). But, despite the different data frames and research designs, all these studies are consistent in their main findings: first, that misinformation is less prevalent, in terms of volume, than the amount of public attention to this problem would suggest; second, that key demographic variables like ideology and age seem to be driving the sharing of misinformation on social media; and third that, on average, people can accurately distinguish between lower- and higher-quality sources but flagging false content is not very effective in reducing misperceptions.

There is a fourth category of studies that use experimental designs outside the natural environment of media platforms, created solely for research purposes, that is, studies that rely on stylized environments that aim to emulate specific features of real platforms (e.g. Jennings and Stroud, 2023; Porter et al., 2018; Rhodes, 2022; Thorson, 2016). This past research tests the effects of different correction strategies on misperceptions and evaluates alternative approaches to content labeling, uncovering important psychological and platform-design mechanisms that can amplify the impact of misinformation. But these studies are also limited by concerns about the external validity of their results, which do not always line up with evidence from experiments conducted in the field (e.g. Aslett et al., 2022). Observational data collected from natural environments, on the other hand, are correlational and unable to identify causal effects. In addition, most past observational research has focused on sharing behavior, which is an imperfect proxy for actual exposure to news, especially if we consider that social media often act as mediators that refer traffic to the web, that is, the web (Lazer et al., 2021). Measuring who retweets what on Twitter, or who shares posts on Facebook, in other words, does not necessarily measure who is seeing which content.

Beyond limitations about measurement and research designs, past work also leaves open theoretical questions about sources of heterogeneity in exposure to misinformation,

especially in terms of which subpopulations are at higher risk of encountering it and engaging with it (Freelon and Wells, 2020). Age and ideology emerge from past work as two of the main predictors of sharing misinformation: older people and people that identify as Republican or Conservative are more likely to engage with unreliable content. However, it is unclear whether these divides also hold for other forms of exposure on platforms other than social media. For instance, incidental exposure is less prominent on the web than on social media: most news consumption on the web results from intentional news seeking (Yang et al., 2020). Past research also raises the question of whether measures of party affiliation (Republican vs Democrat) are enough to characterize the actual ideological diversity of news diets, or the degree of heterogeneity in the news consumed by individuals that share the same party affiliation. Party affiliation captures political identity; the ideological diversity of news diets captures actual behavior and revealed content preferences.

Our analyses contribute new evidence to address some of these questions. We analyze observed exposure to misinformation on the web with two objectives. The first is to characterize the demographic profile of the group of people that consume misinformation. Specifically, we ask: are the demographic divides in exposure to misinformation consistent with what past research has observed on social media? One key difference between our research and past work is that the latter predominantly uses measures of engagement (e.g. shares) rather than measures of exposure (i.e. actual impressions or views). In addition, the web creates a very different news ecosystem because social and algorithmic curation are less prominent in determining exposure to content. Even if some of the traffic is driven by social media referrals, the web is a platform that relies on more intentional news seeking and is less impacted by the network effects of social media. The second objective is to determine how the population consuming unreliable content are engaging other news content: are they doing so at the expense of more reliable news? What are the characteristics of their broader news diets? And are they only reading ideologically congruent content? The answer to these questions is particularly important to understand the broader impacts of misinformation and how to successfully address it with systemic interventions.

High levels of political interest are associated with higher levels of education, so if people consuming misinformation rank also high in their level of interest, designing programs to increase digital literacy is unlikely to have much impact: the main audiences of unreliable content already have the skills. Likewise, the effects of misinformation are likely to vary across people with different levels of ideological diversity in their news diets. Survey experiments suggest that cross-cutting exposure is one of the mechanisms that can strengthen individuals' ability to be critical in their evaluation of news (Rhodes, 2022). But if people consuming most misinformation are also those with the most diverse diets, in ideological terms, then encouraging more cross-cutting exposure is unlikely to have the theorized effect.

These, of course, are all empirical possibilities that need to be substantiated with actual data from observed behavior. In the following section, we introduce the data we use to discriminate between these possible scenarios and reconstruct the news behavior of those who consume unreliable content. We propose a novel measure to capture ideological diversity in news diets, which we use in conjunction with measures of overall

interest in news content and demographic correlates to predict exposure to unreliable information. The following section gives more details on our data and methods, including how we operationalize our main dependent variable: exposure to unreliable content. Unlike most research on social media (which measures engagement rather than exposure), we measure exposure to unreliable content both as the count of unreliable pages visited and as the time spent on those pages.

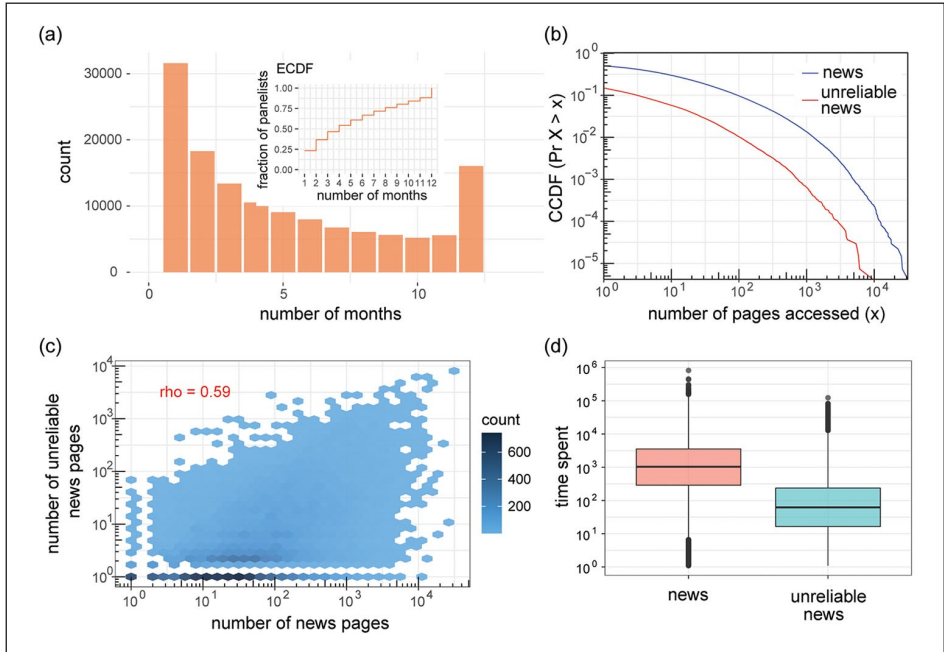
## Data and methods

### Panel data

Our data track web browsing behavior for  $N \sim 140,000$  unique users in the United States for the period of January–December 2018. The data are provided by Nielsen, the media measurement company, and it is weighted to be representative of the US population (weights are also provided by Nielsen). One of the strengths of this data source is that it offers URL-level granularity, which means that we can analyze exposure to all the web pages the panelists visited during the observation window, as well as the time (in seconds) they spent visiting those pages and the parent domains. This data source also offers individual-level demographic attributes, such as gender, age, education, race, ethnicity, and income, covering more demographic dimensions than prior studies. Another strength of this data source is that web logs offer repeated observations for the same individuals, which allows us to control for intra-person temporal variability. As we show in Figure 1(a), half of the panelists in our data are active for at least 4 months, with  $N \sim 16,000$  panelists active for the full 12 months we analyze. Compared with other data sources more restricted in sample size and temporal resolution, our data allow us to model individual exposure to unreliable content controlling for longitudinal trends and both inter- and intra-person variability. In other words, the analysis of these data allows us to detect and measure statistical effects that other approaches, based on aggregated time series or cross-sectional data, cannot measure.

### Variables

**News exposure.** Our main variable of interest is exposure to unreliable content. We identify unreliable sources through the merging of two lists. The first is the list of domains classified as misinformation by Grinberg et al. (2019). This list, which includes  $N=510$  domains, was itself compiled using existing lists produced by journalistic outlets (i.e. BuzzFeed, Politifact, FactCheck.org, Snopes.com) and by prior research (Allcott and Gentzkow, 2017; Guess et al., 2020b). Only 199 of these 510 domains (39%) were visited by our panelists at least once. The second list is provided by NewsGuard, a journalism and technology organization that rates the credibility of news and information websites. Each site receives a trust score on a 0–100 scale based on nine criteria, five related to credibility (i.e. the site does not repeatedly publish false content; gathers and presents information responsibly; regularly corrects or clarifies errors; handles the difference between news and opinion responsibly; avoids deceptive headlines) and four related to transparency (i.e. the website discloses ownership and financing; clearly labels



**Figure 1.** Description of the data. (a) Longevity of panelists measured as the number of months in which they are active in the data (inset shows the empirical cumulative distribution function). (b) Complementary cumulative distribution function of the number of news pages (blue curve) and unreliable news pages (red curve) visited by all panelists. One percent of the panelists are responsible for visiting 43.2% of news pages but, for unreliable news, 1% of the panelists are responsible for visiting 65.3% of the pages. Exposure to unreliable content is thus more concentrated/skewed than exposure to news content. (c) Association between the number of news and unreliable news pages among panelists that visited both types of sites. (d) Distribution of time spent consuming news and unreliable news among panelists that visited both types of sites.

advertising; reveals who is in charge, including possible conflicts of interest; and the site provides names of content creators, along with either contact or biographical information). Websites with a score  $< 60$  points receive a red rating, which means that they are deemed to generally fail to meet basic standards of credibility and transparency and can be considered as unreliable sources, consistent with prior research (e.g. Aslett et al., 2022). Among the 69 misinformation outlets that are identified by Grinberg et al. (2019), visited by our panelists, and rated by NewsGuard, 58 of them (84%) have credibility scores  $< 60$ , signaling high recall rate and correlation between the two lists. We add these websites to our list of unreliable sources using the scores produced in 2018 (the same year for which we have the panel data). This brings the total of news domains classified as unreliable that were visited at least once by our panelists to  $N = 504$ . For each panelist we then count the number of pages from unreliable sources visited in each month, as well as the time spent on those pages (in seconds).



In addition to exposure to unreliable content, we also measure exposure to news. To identify web domains that classify as reliable news (separate from the unreliable sources discussed in the previous paragraph), we merged the lists of news sources used in five previous published studies (i.e. Bakshy et al., 2015; Budak et al., 2016; Grinberg et al., 2019; Peterson et al., 2021; Yang et al., 2020). This merged list has a total of  $N=813$  domains, but only  $N=707$  (87%) was visited at least once by our panelists. In Figure 1(b), we show the cumulative distribution function (CCDF) for the number of pages visited classified as news (blue curve) and those classified as unreliable news (red curve). These functions allow us to determine what percentage of the panelists visit at least  $x$  number of pages in each category. In line with prior research, the curves show that  $\sim 41\%$  of the panelists do not visit any news page (which means that a large fraction of the online population opts out of news, Yang et al., 2020); and that the vast majority, that is  $\sim 79\%$ , do not visit any page classified as unreliable (again, confirming the finding that misinformation affects a small fraction of the online population, e.g. Allen et al., 2020). Still,  $N \sim 1400$  panelists (about 1%) visit at least 100 unreliable news pages during our observation period. This 1% of the panelists are responsible for visiting 65.3% of all unreliable pages in our data; by contrast, 1% of the panelists make up for 43.2% of all the reliable news pages visited during our observation window. Exposure to unreliable content is thus more concentrated and skewed than exposure to reliable news content.

These two variables — the count of news pages and the count of unreliable pages that panelists visit—are moderately correlated, as we show in Figure 1(c). As mentioned, few panelists visit many news pages, and even fewer visit many unreliable sources; but for those who have higher counts, there is clearly a positive association, which offers the first piece of evidence supporting what we call the puzzle of misinformation. One of the main questions we want to address is whether this association remains statistically significant once we control for demographic covariates known to predict news consumption and engagement with unreliable content. Finally, in addition to measuring individual-level engagement with the news using the count of pages accessed, we also keep track of the time spent on those pages. In Figure 1(d) we plot the distribution of the average time panelists spent reading news and unreliable sources (time spent is measured in seconds and is aggregated for the full year in this figure). In line with what the CCDF curves in panel 1B show, there is also less engagement with unreliable news compared with reliable sources when engagement is measured as time spent on those domains.

*Ideological scores of news domains.* To measure the ideological diversity of news diets, we first assign a media bias label to news domains based on the labels provided by Ad Fontes, which samples prominently featured articles on news domains' websites and employs an ideologically balanced panel of experts to rate each news article's ideological slant. These ratings, which have been used in past work (e.g. Aslett et al., 2022), range from  $-38.5$  (most extreme liberal bias) to  $+38.5$  (most extreme conservative bias). In the online Supplemental Appendix, we report a robustness check employing audience-based metrics of ideological scores also used in past research (e.g. Tyler et al., 2021; Yang et al., 2020). Our findings remain robust with both sets of ratings.

*Demographic variables.* In addition to behavioral indicators of exposure to content on the web, we also have demographic information about the panelists: we have data about their



**Table 1.** Demographic variables.

Variables	%	
	Weighted	Unweighted
Gender		
Male	42	35
Female	58	65
Age		
18–24	10	14
25–34	18	21
35–54	36	35
55+	36	31
Education		
High school or less	25	22
Some college	41	42
College graduate	22	23
Postgraduate	12	12
Employment		
Working	72	35
Not working	28	65
Race		
White	76	74
Black	13	16
Asian	4	4
Hispanic origin		
Hispanic	14	10
Not Hispanic	86	90
Income		
<24,999	11	23
25,000–49,999	23	28
50,000–74,999	24	21
75,000–99,999	17	13
100,000–149,999	15	10
>150,000	9	5
<i>N</i>	136,054	136,054

Percentage of panelists in each demographic group, according to weighted and unweighted data. All statistical analyses use the weighted data. *N* refers to the number of panelists.

gender, age, education, employment status, race, ethnicity, and income. Table 1 offers details of the proportion of panelists that fall in each demographic category, estimated with and without the weights provided by Nielsen (the statistical analyses that follow use the weighted data). Prior research on online misinformation has highlighted the positive impact that age has on sharing false content (e.g. Guess et al., 2019), but prior research has also shown that other attributes, like education and gender are, in general, also important predictors to understand who consumes news and political content (i.e. Mak,

2023; Scharkow et al., 2020; Shehata and Strömbäck, 2011). For instance, this past work shows that there is a clear gender divide: compared with men, women consume less news. Whether these gaps vary depending on the channel used to access news (i.e. social media vs the web) is an empirical question. Here, we focus on these divides as they appear in exposure to unreliable content on the web.

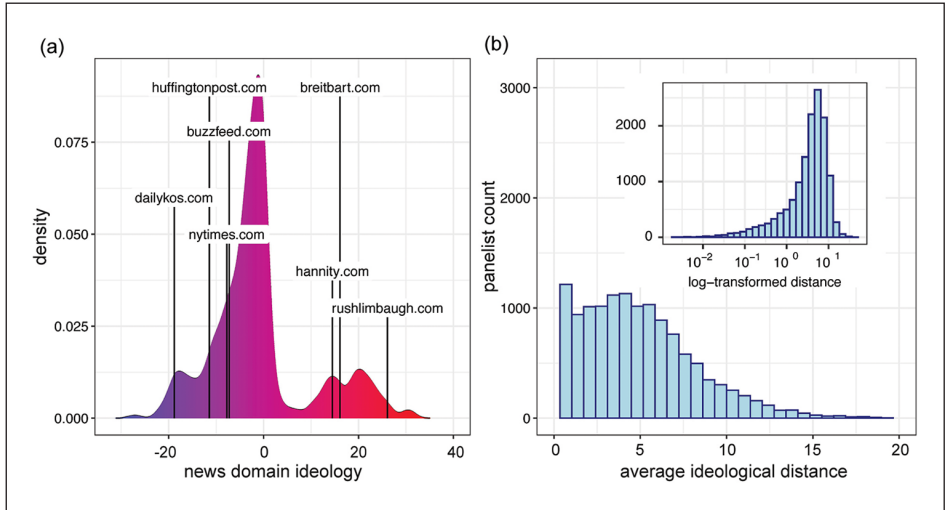
### *Modeling approach*

The observations in our panel data are nested: news sessions are nested within individual panelists who are nested within temporal aggregations (in our case, months). This data structure is rich enough to allow us to model intra- and inter-person variability. One common assumption when applying statistical models to observational data is that individuals with different characteristics behave differently; but it is also true that the same individuals may exhibit different news-seeking behaviors over time. Events exogenous to the data (i.e. political affairs) may also drive overall levels of interest in the news. We account for all these sources of variability using linear mixed-effects models (Bates et al., 2015; Gelman and Hill, 2007). We use panelist ID and month as random effects and the demographic variables described in the previous section as fixed effects (again, using weights). Our main output variable is exposure to misinformation, which we make operational using the number of unreliable pages accessed and the time spent on those pages.

Our two main explanatory variables are interest in the news, which we measure as the number of news pages accessed; and the ideological diversity of news exposure, which we measure as the average distance in the ideological slants of news pages visited. The ideological score of the news page was the score of its news domain. Compared with prior research, we have more controls and more granular measurements, especially when it comes to measuring exposure to misinformation: instead of sharing behavior, we track actual exposure in the form of pages accessed and time spent on those pages. And because of the size of our panel data, which is orders of magnitude larger than the sample sizes used in most prior work, we can also model rare behavior (i.e. exposure to misinformation) with more precision. Our panel data also allow us to leverage the fixed-effects model to investigate intra-person changes, as we detail below and in our online Supplemental Appendix.

## **Results**

The first step in our analyses is to measure the diversity of news diets. In Figure 2(a) we show the distribution of ideological scores for the news domains in our data, with a small subset of labels shown for illustration, i.e. The Daily Kos ([dailykos.com](http://dailykos.com)) is labeled as having a liberal slant; The Rush Limbaugh Show ([rushlimbaugh.com](http://rushlimbaugh.com)) is labeled as having a conservative slant. Using these scores, we calculate the pairwise ideological distance for all news pages visited by the panelists, and then average those distances to assign each panelist a monthly score that we use as a measure of the diversity of their news diet. For instance, if panelist  $i$  accessed three pages in month  $m$  from the Activist Post, BuzzFeed, and the New York Times, we first calculate the ideological distance between each pair of these outlets, and we then average the pairwise distances to a summary statistic for panelist  $i$  in month  $m$ . This average is smaller for panelists that access

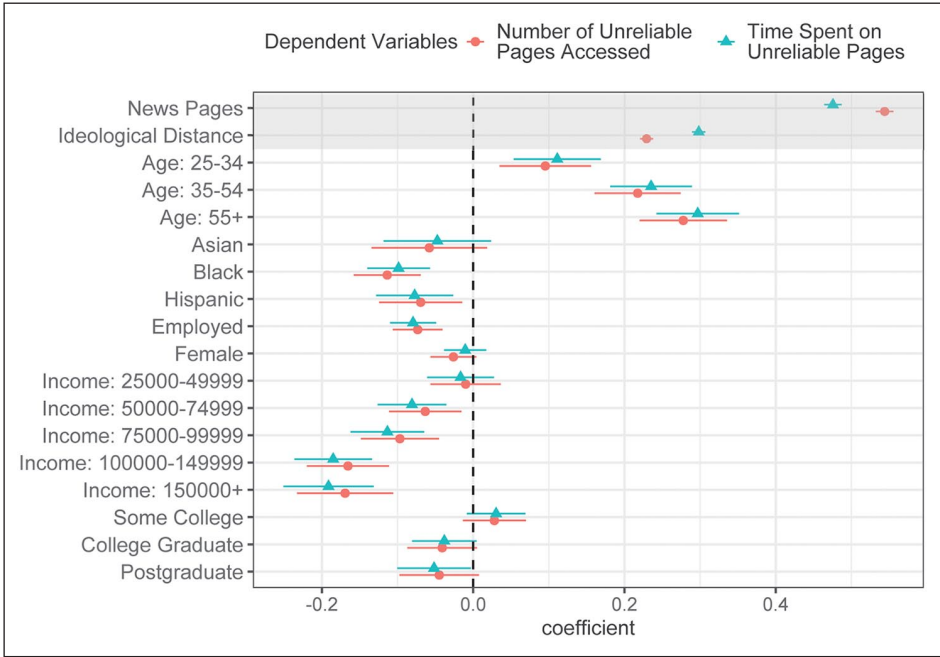


**Figure 2.** Ideological diversity of news diets. (a) Ideological scores of the news domains visited by the panelists. (b) Ideological diversity of panelists’ news diets, measured as the average pairwise distance between the ideology of the news pages they visited (inset shows the log-transformed version of the measure).

ideologically similar domains—so the larger it is in magnitude, the more diverse we can consider their news diet to be in a given month. In Figure 2(b) we show the distribution of these diversity scores (the inset shows the log-transformed version of the measure). In general, many panelists have homogeneous news diets, but there is also heterogeneity: because of the way in which the ideology scores are calculated, scores that seem small in magnitude are still signaling meaningful diversity in the content accessed by the panelists. This measure allows us to differentiate panelists that would otherwise look identical if we were using their self-disclosed ideology.

The ideological composition of news diets and the overall levels of engagement with news are our two key variables. As we show in Figure 3, the most important predictor of exposure to unreliable content is overall interest in the news (as measured by the number of news pages accessed, log-transformed); controlling for this, panelists with news diets that show higher levels of ideological diversity are also exposed to more misinformation (note that the variables have been normalized using the mean value and standard deviation, so they are all on the same scale). As expected, given prior research, the most important demographic covariate is age: older panelists clearly consume more misinformation than their younger counterparts. But within this age category, people with higher levels of interest and more ideologically diverse diets are consuming even more unreliable content.

In the Supplemental Appendix, we additionally show that this puzzle holds when we control the partisan leaning of panelists (Figure A1) or when we divide our panelists into five quantiles based on their partisan leaning (Table A3 and Table A4). In other words, for liberals and conservatives alike, ideologically diverse news diets consistently predict more exposure to unreliable content. Table A2 adopts fixed-effects models to control



**Figure 3.** Predictors of exposure to unreliable content.

These estimates result from a linear mixed-effects model with panelist ID and month as random effects. Variables have been scaled (i.e. normalized using the mean value and standard deviation) and log-transformed where appropriate. See Table A1 and A2 in the Supplemental Appendix for additional specifications and details.

time-invariant demographic variables and event shocks. These models show that the puzzle holds when we examine intra-person effects, that is, if the same person’s news diet becomes more diverse from one month to the next, their exposure to unreliable content also becomes higher. Section B in the Supplemental Appendix reports another set of robustness checks using an alternative measure of ideology, yielding similar findings.

Table A1 and Table B1 in the Supplemental Appendix summarize outputs of alternative models with different specifications. Higher education levels, for instance, are associated with higher exposure to misinformation, but the association disappears when overall interest in news and the diversity of news diets are added to the model. Women, Asian, Black, and Hispanic panelists are also consuming less misinformation than White, male panelists. This is also consistent with what prior research shows in the context of social media (i.e. Guess et al., 2019). Overall, our results suggest that misinformation is not crowding out more reliable sources; instead, panelists with larger and more diverse news diets are also more likely to access unreliable content.

## Discussion

Recent research has established that misinformation amounts to a small fraction of all the information circulating online, and that only a small number of people engage actively

with it. Our findings are consistent with this pattern: only 1% of our panelists visit at least 100 unreliable pages, and they spend less time on those sources than on reliable news. But our findings also illuminate an important pattern not discussed in prior work: it is people with higher levels of news exposure *and* more diverse news diets that are also more likely to access unreliable sources. According to these findings, the driver behind exposure to misinformation is overall political interest, which results in more expansive and diverse news consumption.

We know that political interest is correlated with education, and we also know that this is one of the main factors associated with political involvement: people exhibiting more intense news-seeking behavior are also more likely to be deeply involved in politics (Krupnikov and Ryan, 2022). This group of people are far from representing the population at large but they have a disproportionate influence in the distribution of political attention. The fact that they are also the main consumers of misinformation has consequences for how we think about effects and interventions. Most of the existing literature evaluating interventions to counteract misinformation predominantly focus on psychological processes and individual skills, such as levels of literacy (e.g. Guess et al., 2020a; Vraga et al., 2022), accuracy prompts (e.g. Pennycook et al., 2021), and fact-checking (e.g. Walter et al., 2020). Our results suggest that those who encounter most unreliable content (on the web, in the United States) are also those with broader, more diverse news diets: they have literacy skills and access to reliable and accurate information. Seen through this light, the problem of misinformation is not a problem of literacy or correction. It is more systemic, and it requires contextualizing exposure to unreliable content within the larger landscape of evolving news habits and the choices people make (often prompted by technological affordances and algorithmic curation) when navigating the information environment.

Our results are constrained by limitations that affect the findings' generalizability to other contexts. The web is an information ecosystem on its own, different from other widely studied contexts such as social media newsfeeds. Exposure to unreliable content on social media may be more incidental and therefore less driven by political interest. There is very little research measuring exposure on social media platforms (recent exceptions are González-Bailón et al., 2023, Guess et al., 2023, Nyhan et al., 2023). Most studies of social media focus on engagement (i.e. sharing or commenting), a measure that undercounts the actual number of people exposed. It is likely that political interest still plays a role in the posting of unreliable content: higher levels of involvement in politics are, after all, associated with higher activity on social media (Krupnikov and Ryan, 2022). But given current measurement limitations, we cannot compare exposure to unreliable content on the web with exposure to unreliable content on social media. In addition, our estimates are based on desktop-only web browsing. Future research should incorporate exposure through mobile devices. Another generalizability concern rises from our strategy of identifying misinformation exposure using URL visits, while social media newsfeeds (e.g. on Facebook and Twitter) and other digital media platforms (e.g. Reddit, Instagram, YouTube, TikTok, and WhatsApp groups) contain misinformation in other modalities. For example, Yang et al. (2023), Peng et al. (2023) and Brennen et al. (2021) highlight the prevalence of visual misinformation in the forms of memes, videos, and photographs. These types of misinformation are more casual and light-hearted, which

may attract audiences with lower levels of political interest. Thus, visual misinformation may reduce or reverse the positive association we identify here between political interest, news diet diversity, and exposure to unreliable content.

A second limitation is that we do not consider effects. Our analyses are observational, and they center on identifying patterns in news seeking. We can determine whether misinformation is consumed at the expense of reliable news (it is not); whether those exposed to more unreliable content have news diets that are ideologically siloed (they do not); and which groups of people are most likely to consume unreliable content (older white males with higher levels of political interest). However, on their own, these analyses cannot tell us whether exposure to unreliable content has any other behavioral or psychological effects—for instance, a positive impact on other forms of political engagement. It is the possibility of these effects that turns the unrepresentative group of avid news consumers into the opinion leaders that could serve as credibility assessors (Wagner and Boczkowski, 2019); but also into the brokers that could pass misinformation to broader audiences through other, offline channels. Future research needs to evaluate this possibility, which will require linking observed exposure across media channels, interpersonal networks, and the impact of exposure on beliefs, attitudes and behaviors. Our analyses are also agnostic about the persuasive impact of unreliable content, especially given that misinformation's most likely audiences are also the better informed. The exposure measures we analyze do not offer any insights about perceptions or attitudes. As qualitative research shows, consuming news is not necessarily an indicator of perceived credibility or impact on political beliefs (Munyaka et al., 2022; Weeks et al., 2021). Reading and sharing news (which is what we can measure with observational data) does not, on its own, tell us how much political learning or updating is happening as a consequence of that exposure.

Third, our results are suggestive of the nested paradoxes at the heart of how misinformation finds an audience. Those with higher levels of news consumption are also those more likely to spend more time on unreliable news; and if, as past research suggests, more intense information seeking is correlated with deeper political involvement, then this sector of the population is also more likely to spread that content through their discussion networks, online and offline, thus amplifying its reach. Whether this amplification is accompanied with corrective statements or not is an empirical question that, again, our data cannot illuminate; but it is possible that the better informed are in fact contributing to give more visibility (and credibility) to unreliable content. This possibility is supported by the fact that political engagement often serves as a key antecedent of misinformation sharing (Valenzuela et al., 2019). Unpacking these paradoxes is, therefore, central to determine the best point of entry to implement interventions designed to curtail the spread of misinformation.

And fourth, as emphasized throughout the study, the findings reported here are bounded by the regional context of our data (i.e. the United States during a specific 12-month period). Considering the high level of partisan sorting, political polarization, and asymmetrical misinformation supply that characterize the US media ecosystem, we are agnostic about the generalizability of our findings to other contexts and the universality of the puzzle we discuss. Our findings suggest the problem of misinformation requires a more systemic approach than what individual-oriented studies, focused on psychological mechanisms, can afford. Different media and political systems can drastically change the



information-seeking behavior of news consumers. As web tracking panel data become more accessible in other political contexts, scholars should try to replicate our study to determine whether the puzzle appears in other political contexts as well.

Ultimately, our study sheds lights on the normative tensions underlying epistemic theories of democracy and how we should think about the “informed citizenry” in the current media environment. Healthy news habits involve, in expectation, (1) consuming more news to keep up with current affairs; (2) contrasting ideologically diverse sources; and (3) limiting exposure to unreliable information. And yet, our results suggest that when (1) and (2) are in place, (3) seems more difficult to achieve. These are behaviors that affect only a small fraction of the population, that is, the group of people with high political interest and high levels of news engagement. Yet it begs the normative question of how to encourage news consumption while shielding users from unreliable sources—content they are more likely to see when their news diets become more intense and diverse. The answer to this question requires thinking about the overall information architecture and how people navigate it as a function of their choices and the opportunities the architecture offers.

### Authors' note

All authors have agreed to the submission and that the article is not currently being considered for publication by any other print or electronic journal.


### Acknowledgements

We are grateful to Harmony Labs for facilitating access to the data and for crucial research support.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Work on this article was partly funded by NSF Grant 2017655.

### ORCID iDs

Alvin Zhou  <https://orcid.org/0000-0001-5410-9712>

Tian Yang  <https://orcid.org/0000-0003-2642-4760>

### Supplemental material

Supplemental material for this article is available online.

### References

- Allcott H and Gentzkow M (2017) Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31(2): 211–235. Available at: <http://www.jstor.org/stable/44235006>
- Allen J, Howland B, Mobius M, et al. (2020) Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances* 6(14): eaay3539.
- Altay S, Nielsen RK and Fletcher R (2022) Quantifying the “infodemic”: people turned to trustworthy news outlets during the 2020 coronavirus pandemic. *Journal of Quantitative Description: Digital Media* 2: 1–29.

- Aslett K, Guess AM, Bonneau R, et al. (2022) News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions. *Science Advances* 8(18): eabl3844.
- Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239): 1130–1132.
- Bates D, Mächler M, Bolker B, et al. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1–48.
- Benkler Y, Faris R and Roberts H (2018) *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford: Oxford University Press. Available at: <https://books.google.com/books?id=6hhnDwAAQBAJ>
- Brennen JS, Simon FM and Nielsen RK (2021) Beyond (mis)representation: visuals in COVID-19 misinformation. *The International Journal of Press/Politics* 26(1): 277–299.
- Budak C, Goel S and Rao JM (2016) Fair and balanced? Quantifying media bias through crowd-sourced content analysis. *Public Opinion Quarterly* 80(Suppl. 1): 250–271.
- Chen Y, Conroy NJ and Rubin VL (2015) Misleading online content: recognizing clickbait as “False News.” In: *Proceedings of the 2015 ACM on workshop on multimodal deception detection*, Seattle, WA, 13 November, pp. 15–19. New York: ACM.
- Eady G, Paskhalis T, Zilinsky J, et al. (2023) Exposure to the Russian Internet research agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior. *Nature Communications* 14(1): 62.
- Fletcher R, Kalogeropoulos A and Nielsen RK (2023) More diverse, more politically varied: how social media, search engines and aggregators shape news repertoires in the United Kingdom. *New Media & Society* 25(8): 2118–2139.
- Freelon D and Wells C (2020) Disinformation as political communication. *Political Communication* 37(2): 145–156.
- Gelman A and Hill J (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Grinberg N, Joseph K, Friedland L, et al. (2019) Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363(6425): 374–378.
- Guay B, Berinsky A, Pennycook G, et al. (2022) Examining partisan asymmetries in fake news sharing and the efficacy of accuracy prompt interventions. *PsyArXiv*. Epub ahead of print 14 April. DOI: 10.31234/osf.io/y762k.
- Guess A, Aslett K, Tucker J, et al. (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: 1–48.
- Guess A, Nagler J and Tucker J (2019) Less than you think: prevalence and predictors of fake news dissemination on Facebook. *Science Advances* 5(1): eaau4586.
- Guess AM, Lerner M, Lyons B, et al. (2020a) A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences* 117(27): 15536–15545.
- Guess AM, Nyhan B and Reifler J (2020b) Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour* 4(5): 472–480.
- González-Bailón S, Lazer D, Barberá P, et al. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, 381(6656), 392–398. <https://doi.org/doi:10.1126/science.ade7138>
- Guess AM, Malhotra N, Pan, J et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656), 398–404. <https://doi.org/doi:10.1126/science.abp9364>
- Huszár F, Ktena SI, O’Brien C, et al. (2022) Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences* 119(1): e2025334119.

- Jennings J and Stroud NJ (2023) Asymmetric adjustment: partisanship and correcting misinformation on Facebook. *New Media & Society* 25(7): 1501–1521.
- Krupnikov Y and Ryan JB (2022) *The Other Divide: Polarization and Disengagement in American Politics*. Cambridge: Cambridge University Press.
- Lazer D, Baum MA, Benkler Y, et al. (2018) The science of fake news. *Science* 359(6380): 1094–1096.
- Lazer D, Hargittai E, Freelon D, et al. (2021) Meaningful measures of human society in the twenty-first century. *Nature* 595(7866): 189–196.
- Mak MKF (2023) Explaining the gender gap in news access across thirty countries: resources, gender-bias signals, and societal development. *The International Journal of Press/Politics* 28(3): 533–557.
- Moretto M, Ortellado P, Kessler G, et al. (2022) People are more engaged on Facebook as they get older, especially in politics: evidence from users in 46 countries. *Journal of Quantitative Description: Digital Media* 2: 1–20.
- Munyaka I, Hargittai E and 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). jots.v1i4.62.
- Nelson JL and Taneja H (2018) The small, disloyal fake news audience: the role of audience availability in fake news consumption. *New Media & Society* 20(10): 3720–3737.
- Nyhan B, Settle J, Thorson E, et al. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*. <https://doi.org/10.1038/s41586-023-06297-w>
- Nielsen RK and Fletcher R (2022) Concentration of online news traffic and publishers' reliance on platform referrals: evidence from passive tracking data in the UK. *Journal of Quantitative Description: Digital Media* 2: 1–23.
- Peng Y, Lu Y and Shen C (2023) An agenda for studying credibility perceptions of visual misinformation. *Political Communication* 40(2): 225–237.
- Pennycook G and Rand DG (2019) Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116(7): 2521–2526.
- Pennycook G and Rand DG (2021) Research note: examining false beliefs about voter fraud in the wake of the 2020 presidential election. *Harvard Kennedy School Misinformation Review*. Epub ahead of print 11 January. DOI: 10.37016/mr-2020-51.
- Pennycook G, Epstein Z, Mosleh M, et al. (2021) Shifting attention to accuracy can reduce misinformation online. *Nature* 592(7855): 590–595.
- Peterson E, Goel S and Iyengar S (2021) Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign. *Political Science Research and Methods* 9(2): 242–258.
- Porter E, Wood TJ and Kirby D (2018) Sex trafficking, Russian infiltration, birth certificates, and pedophilia: a survey experiment correcting fake news. *Journal of Experimental Political Science* 5(2): 159–164.
- Rhodes SC (2022) Filter bubbles, echo chambers, and fake news: how social media conditions individuals to be less critical of political misinformation. *Political Communication* 39(1): 1–22.
- Rossini P, Stromer-Galley J, Baptista EA, et al. (2021) Dysfunctional information sharing on WhatsApp and Facebook: the role of political talk, cross-cutting exposure and social corrections. *New Media & Society* 23(8): 2430–2451.
- Scharkow M, Mangold F, Stier S, et al. (2020) How social network sites and other online intermediaries increase exposure to news. *Proceedings of the National Academy of Sciences* 117(6): 2761–2763.
- Shehata A and Strömbäck J (2011) A matter of context: a comparative study of media environments and news consumption gaps in Europe. *Political Communication* 28(1): 110–134.

- Thorson E (2016) Belief echoes: the persistent effects of corrected misinformation. *Political Communication* 33(3): 460–480.
- Tyler M, Grimmer J and Iyengar S (2021) Partisan enclaves and information bazaars: mapping selective exposure to online news. *The Journal of Politics* 84(2): 1057–1073.
- Valenzuela S, Halpern D, Katz JE, et al. (2019) The paradox of participation versus misinformation: social media, political engagement, and the spread of misinformation. *Digital Journalism* 7(6): 802–823.
- Vosoughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science* 359(6380): 1146–1151.
- Vraga EK, Bode L and Tully M (2022) Creating news literacy messages to enhance expert corrections of misinformation on Twitter. *Communication Research* 49(2): 245–267.
- Wagner MC and Boczkowski PJ (2019) The reception of fake news: the interpretations and practices that shape the consumption of perceived misinformation. *Digital Journalism* 7(7): 870–885.
- Walter N, Cohen J, Holbert RL, et al. (2020) Fact-checking: a meta-analysis of what works and for whom. *Political Communication* 37(3): 350–375.
- Watts DJ, Rothschild DM and Mobius M (2021) Measuring the news and its impact on democracy. *Proceedings of the National Academy of Sciences* 118(15): e1912443118.
- Weeks BE, Menchen-Trevino E, Calabrese C, et al. (2021) Partisan media, untrustworthy news sites, and political misperceptions. *New Media & Society*. Epub ahead of print 28 July. DOI: 10.1177/14614448211033300.
- Yang T, Majó-Vázquez S, Nielsen RK, et al. (2020) Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences* 117(46): 28678–28683.
- Yang Y, Davis T and Hindman M (2023) Visual misinformation on Facebook. *Journal of Communication* 74(4): 316–328.
- Zarocostas J (2020) How to fight an infodemic. *The Lancet* 395(10225): 676.

### Author biographies

Alvin Zhou, PhD, is an Assistant Professor at the Hubbard School of Journalism and Mass Communication at the University of Minnesota. His research focuses on computational social science and strategic communication.

Tian Yang, PhD, is an Assistant Professor at the School of Journalism and Communication at the Chinese University of Hong Kong. He studies political communication, computational social science, and Chinese media.

Sandra González-Bailón, PhD, is the Carolyn Marvin Professor of Communication at the Annenberg School at the University of Pennsylvania, and Director of the Center for Information Networks and Democracy (CIND). Her research lies at the intersection of political communication and computational social science.