Article

Users choose to engage with more partisan news than they are exposed to on Google Search

https://doi.org/10.1038/s41586-023-06078-5

Received: 17 February 2022

Accepted: 12 April 2023

Published online: 24 May 2023

Check for updates

Ronald E. Robertson[1,2]✉, Jon Green[2], Damian J. Ruck[2], Katherine Ognyanova[3], Christo Wilson[2,4] & David Lazer[2]

If popular online platforms systematically expose their users to partisan and unreliable news, they could potentially contribute to societal issues such as rising political polarization[1,2]. This concern is central to the 'echo chamber'[3–5] and 'filter bubble'[6,7] debates, which critique the roles that user choice and algorithmic curation play in guiding users to different online information sources[8–10]. These roles can be measured as exposure, defined as the URLs shown to users by online platforms, and engagement, defined as the URLs selected by users. However, owing to the challenges of obtaining ecologically valid exposure data—what real users were shown during their typical platform use—research in this vein typically relies on engagement data[4,8,11–16] or estimates of hypothetical exposure[17–23]. Studies involving ecological exposure have therefore been rare, and largely limited to social media platforms[7,24], leaving open questions about web search engines. To address these gaps, we conducted a two-wave study pairing surveys with ecologically valid measures of both exposure and engagement on Google Search during the 2018 and 2020 US elections. In both waves, we found more identity-congruent and unreliable news sources in participants' engagement choices, both within Google Search and overall, than they were exposed to in their Google Search results. These results indicate that exposure to and engagement with partisan or unreliable news on Google Search are driven not primarily by algorithmic curation but by users' own choices.

The prevalence of partisan and unreliable online news is a topic of continuing concern among policymakers, civil society organizations and academics[5,25,26]. These concerns often centre around the role of online platforms, such as search engines or social media, in directing people to such content, and the societal effects of such guiding[1,2,27]. The theoretical grounding for such concerns generally involves selective exposure—the tendency to choose political information that is congruent with one's existing beliefs[28,29]—and its counterparts: echo chambers[3,4] and filter bubbles[6,7].

In online settings, echo chambers often centre the role of users' choices, including their URL selection, search query formulation or social networking decisions, and how they can create settings "in which most available information conforms to pre-existing attitudes and biases"[5]. By contrast, filter bubbles often centre the role of algorithmic curation, such as the production of a social media feed or a search results page, whereby content "selected by algorithms according to a viewer's previous behaviors" can create a feedback loop that limits exposure to cross-cutting content[7]. Although the definitions of these two concepts vary and overlap, both can be described in terms of user choice and algorithmic curation[8,9], and both predict a similar observable outcome: partisans will see and select a substantial amount of identity-congruent content.

Recent research on partisan and unreliable online news has primarily focused on users' choices within social media platforms[4,11,12] or during their general web browsing[8,13–16], leaving open questions about the role of algorithmic curation more broadly. The importance of studying news on web search engines, among other platforms, is evident from long-standing concerns about the effects of search engines[30–33], and is urgent in the light of several recent findings. For example, survey and digital trace studies have found that web search plays a central role in directing users to online news[13,34–37], qualitative work has documented patterns of unreliable and false information in web search results[38,39] and lab experiments suggest that politically biased search rankings can influence political opinions[40,41]. However, research on user choice and algorithmic curation within web search has been largely limited to algorithm auditing studies, in which what real users might have seen in their search results was estimated using simulated user behaviour (hypothetical exposure)[17–19,21,22], and digital trace studies, in which what real users might have seen was estimated from available click logs (selected exposure)[8,9,13,42,43].

[1]Stanford University, Stanford Internet Observatory, Stanford, CA, USA. [2]Northeastern University, Network Science Institute, Boston, MA, USA. [3]Rutgers University, School of Communication & Information, New Brunswick, NJ, USA. [4]Northeastern University, Khoury College of Computer Sciences, Boston, USA. ✉e-mail: rer@acm.org

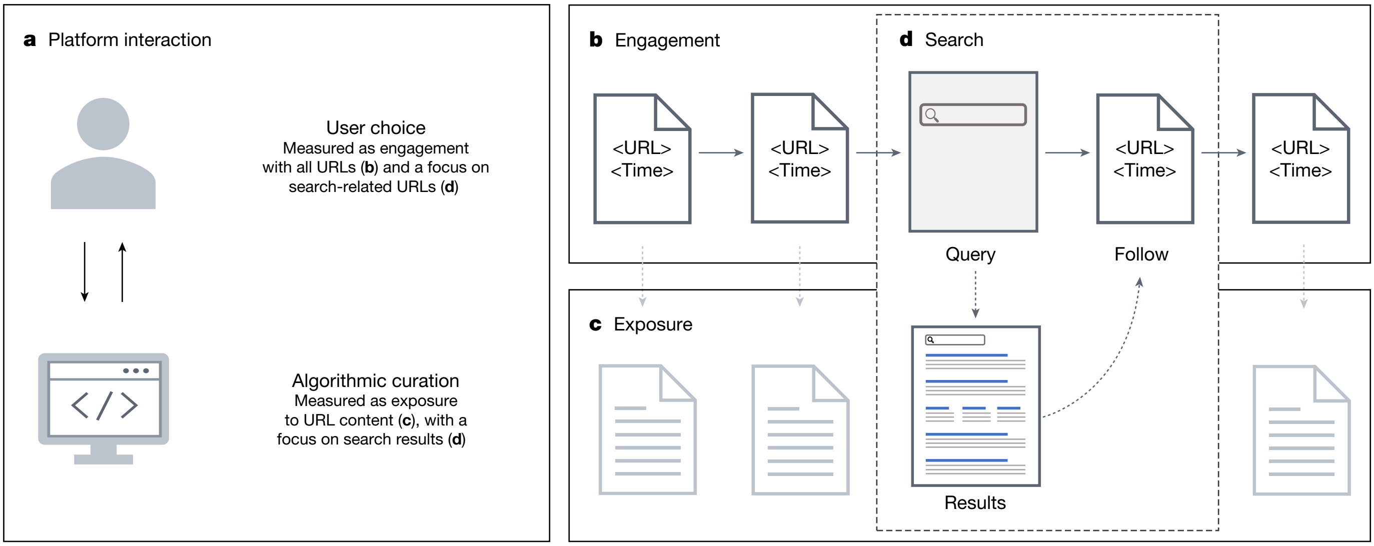


Fig. 1 | Measuring user choice and algorithmic curation as engagement and exposure. a, For each interaction between a user and a platform, we draw a distinction between user choice (top) and algorithmic curation (bottom). **b**, Using our browser extension, we measured user choice as engagement, broadly defined here as a linear log of time-stamped URL visits. **c**, We also used our browser extension to measure algorithmic curation as exposure, broadly defined here as the content that loads in a user's browser window during each URL visit. **d**, We used both exposure and engagement data to examine our participants' interactions with search engines, focusing specifically on exposure to URLs in Google Search results. When our participants engaged with Google Search, we recorded the search query they used and saved the corresponding HTML of the page of search results to which they were exposed. We classified their subsequent engagement with a URL as a search follow if it occurred within 1 minute of the exposure and its domain appeared in the search results (see Methods, 'Exposure and engagement datasets'). Further background information on each measure is available in Methods ('Exposure and engagement definitions').

We operationalize the two sides of a user–platform interaction as exposure and engagement, defining exposure as the URLs people see while visiting a platform and engagement as the URLs that people interact with while on that platform or while browsing the web more broadly (Fig. 1). Although overall engagement provides an aggregate measure of user choice that includes all external influences (including all online and offline factors), we also define follows as the intersection of these data types—engagement conditional on exposure—to isolate the subset of overall engagement choices that were made immediately after exposure to a page of algorithmically curated results. These terms and distinctions build on previous work by incorporating a wider variety of on- and off-platform web behaviours (see Methods, 'Exposure and engagement definitions'). For example, Bakshy et al.[7] used internal Facebook data to measure exposure as cases "in which a link to the content appears on screen in an individual's News Feed," and described on-platform engagement as "clicks" and "consumption," but did not measure off-platform behaviour. So far, studies involving ecologically valid exposure and engagement data—what real users were shown and did during their everyday use of a platform—have largely been limited to internal studies published by social media platforms[7,24].

Here we advance research on user choice and algorithmic curation through a two-wave study in which we deployed a custom web browser extension to collect ecologically valid measures of both exposure and engagement on Google Search. During the 2018 and 2020 US election cycles, we invited participants to complete a survey and voluntarily install our extension with informed consent. We then merged those exposure and engagement data with domain-level (for example, bbc.com) measures of partisan and unreliable news, and used an unsupervised method on the text of participants' queries to quantify differences in their search behaviour. Paired with the surveys, these data enabled us to examine differences among groups with characteristics, such as partisan identification and age, that have previously been linked to greater interaction with partisan or unreliable news[13,23,44].

Results from both study waves show that participants' partisan identification had a small and inconsistent relationship with the amount of partisan and unreliable news they were exposed to on Google Search, and a more consistent relationship with the search results they chose to follow and their overall engagement. Differences in participants' demographic characteristics and the content of their search queries largely explained the small differences we found for exposure to partisan and unreliable news on Google Search by partisan identity, suggesting an absence of filter bubbles in this context. However, the more consistent differences we observed in participants' follows and overall engagement with partisan and unreliable news, representing their on- and off-platform choices, suggest at least some degree of online echo chambers. These findings shed light on the role of Google Search in leading its users to partisan and unreliable news, highlight the importance of measuring both user choice and algorithmic curation when studying online platforms and are consistent with previous work on general web browsing[8,9] and Facebook's News Feed[7].

Collecting exposure and engagement data

From October to December 2018 and from April to December 2020, we recruited participants to take a survey and optionally install a custom browser extension we made for Chrome and Firefox. In the survey, participants self-reported both their seven-point partisan identification (7-point PID; strong Democrat, not very strong Democrat, lean Democrat, independent, lean Republican, not very strong Republican, strong Republican) and their age, which we assigned to one of four age groups (18–24, 25–44, 45–64, 65+). We used a seven-point scale, rather than a three-point one (Democrat, independent, Republican), because strong partisans were oversampled in the 2018 survey and may differ in important respects from respondents with weaker partisan attachments (see Methods, 'Opinion surveys and participant samples').

To measure exposure (the URLs shown to participants; Fig. 1c), we designed our extension to save HTML snapshots of the Google Search results that loaded in participants' web browsers, and used a custom parser to extract the URLs from those snapshots. For follows from Google Search (on-platform URL engagements; Fig. 1d), we collected

Google History in 2018, which tracks activity on Google services and is available at myactivity.google.com, and Tab Activity in 2020, which is an improved measure of engagement that we built to monitor the active browser tab. We did not use Google History to measure follows in 2020 because of changes in the accessibility of those data. Tab Activity enabled us to more directly link exposures to engagements, which we used not only to identify follows as URL visits that occurred immediately after and within 60 s of an exposure (as we did with Google History in 2018), but also to check whether the followed domain matched any of the search results in that exposure. To measure overall engagement (all URLs engaged with; Fig. 1b), we collected Browser History data; Browser History is an application programming interface (API) that is built in to Firefox and Chrome and provides users with an interface for viewing their past website visits. The third-party data collections, Browser History and Google History, include data before the installation of our extension, but our custom data collections, the HTML snapshots and Tab Activity, collected data only while the extension was installed (see Methods, 'Exposure and engagement datasets').

In 2018, we collected exposure to 102,114 Google Search result pages for 275 participants, follows on 279,680 Google Search results for 262 participants and overall engagement with 14,677,297 URLs for 333 participants. In 2020, we collected exposure to 226,035 Google Search result pages for 459 participants, follows on 69,023 Google Search results for 418 participants and overall engagement with 31,202,830 URLs for 688 participants (Extended Data Table 1). Most participants' web searches were conducted on Google Search in both waves (74.2% in 2018, 68.6% in 2020; Supplementary Information Table 4), rather than on other popular search engines, such as Bing and Yahoo.

Measuring web domains and search queries

Following past work[7,13,23,43], we focused on URLs from news domains (Methods, 'Web domain scores and classifications'), which we identified using a compendium of four lists used in research since 2015[7,23,45,46] (see Methods, 'News classifications'). To classify news domains as unreliable, we used a combination of two lists used in previous work (see Methods, 'Unreliable news classifications'). The first is from peer-reviewed work that classified news domains as unreliable on the basis of a manual review of their editorial practices[23]. The second was obtained from NewsGuard, an independent organization that employs journalists and editors to review and rate news domains on the basis of nine journalistic criteria[46].

To score the partisanship of news domains, we used a measure of partisan audience bias derived from the differential sharing patterns of Democrats and Republicans in a large, virtual panel of Twitter users[19]. These scores range from −1 to 1, with a score of −1 indicating that a domain was shared only by Democrats in the Twitter panel, a score of 1 indicating that it was shared only by Republicans in the panel and a score of 0 indicating that an equal proportion of Democrats and Republicans shared it on Twitter. For the Google Search results we collected, we aggregated partisanship scores at the search engine results page (SERP) level by applying a rank-weighted average[19] to the news domains appearing on each SERP. This measure places more weight on the scores of domains that appear near the top of the search rankings, which often receive a disproportionate amount of user attention and number of clicks[10] (see Methods, 'Partisan news scores').

Unlike the news feeds produced by social media platforms, Google Search results pages are produced through a more active information-seeking process that depends on a user-selected search query[47–49]. We therefore measured the content of participants' queries using pivoted text scaling[50], a form of principal component analysis performed on a truncated word co-occurrence matrix to identify orthogonal latent dimensions that explain decreasing shares of variation in the co-occurrence of commonly used words. This unsupervised approach is useful for our application because it does not rely on external sources (such as dictionary-based approaches) and it does not use extra participant-level information (such as partisan identities) in the estimation stage that could risk introducing collinearity into subsequent modelling. It is also appropriate for our use case of characterizing variation in search queries because it was designed specifically for short documents, for which other unsupervised methods such as topic models are less efficient (Methods, 'Search query analysis').

Partisan news

In Fig. 2a,c, we compare the average news partisanship of participants' (1) exposure via Google Search results, (2) follows on Google Search results and (3) overall engagement. Using the Kruskal–Wallis H test to examine differences in partisan news by partisan identity (see Methods, 'Descriptive analysis'), we found significant differences in each study wave and dataset (all $P < 0.01$; Extended Data Table 2). For both study waves, we found that the partisan gap—the difference in news partisanship between the average strong Republican and the average strong Democrat—was small for exposure (0.062 in 2018, 0.037 in 2020), larger for follows (0.106 in 2018, 0.125 in 2020) and largest for overall engagement (0.206 in 2018, 0.134 in 2020).

As in past work, we contextualize these partisan gaps by noting comparable gaps in scores between popular news outlets[51]. For exposure, the partisan gap was comparable to the gap in scores between MSNBC (−0.624) and *Mother Jones* (−0.697) for 2018, and the gap between *The Washington Post* (−0.234) and *The New York Times* (−0.260) for 2020. For follows, the partisan gap for both years was comparable to the gap between Newsmax (0.688) and *InfoWars* (0.782). Last, for overall engagement, the partisan gap was comparable to the gap between *Salon* (−0.593) and *Jacobin* (−0.803) for 2018, and comparable to the gap between Fox News (0.608) and *Breitbart* (0.742) for 2020. These results indicate that, on average, Google Search's algorithmic curation exposes its users to less identity-congruent news than they choose to engage with.

Unreliable news

Examining the average proportion of unreliable news in participants exposure, follows and overall engagement by partisan identity, we found a pattern that was similar to the one observed for partisan news (Fig. 2b,d). Again using the Kruskal–Wallis H test, the only significant differences we found for unreliable news among partisan identities were in participants' overall engagement ($P < 0.01$ in 2018 and 2020; Extended Data Table 2). For the average participant, unreliable news was less prevalent in the URLs they were exposed to on Google Search (2.05% in 2018, 0.72% in 2020) than in their Google Search follows (2.36% in 2018, 0.93% in 2020) or overall engagement (3.03% in 2018, 1.86% in 2020).

For exposure, these percentages represent the fraction of SERPs that contained at least one unreliable news domain (URL-level percentages are available in Supplementary Information Table 6). For engagement, the prevalence of unreliable news was asymmetric across partisan identities, with strong Republicans engaging with or following more unreliable news than strong Democrats, but generally not being exposed to more in their Google Search results (Extended Data Table 2). These results indicate that Google Search's algorithmic curation can be a conduit for exposure to unreliable news, but not to the degree that users' engagement choices are, especially among strong Republicans.

We also found that unreliable news was generally uncommon and concentrated among a small number of participants (Fig. 3). Using the percentage of participants that accounted for 90% of all exposures, follows or engagements as a measure of concentration, we found a pattern similar to the one we found for participant-level averages. Specifically, we found that unreliable news was the least concentrated among a small number of participants for exposure, with 31.3% of participants accounting for 90% of all unreliable news exposures in 2018 (25.1% in 2020), and more concentrated in both follows (28.2% in 2018, 9.8% in

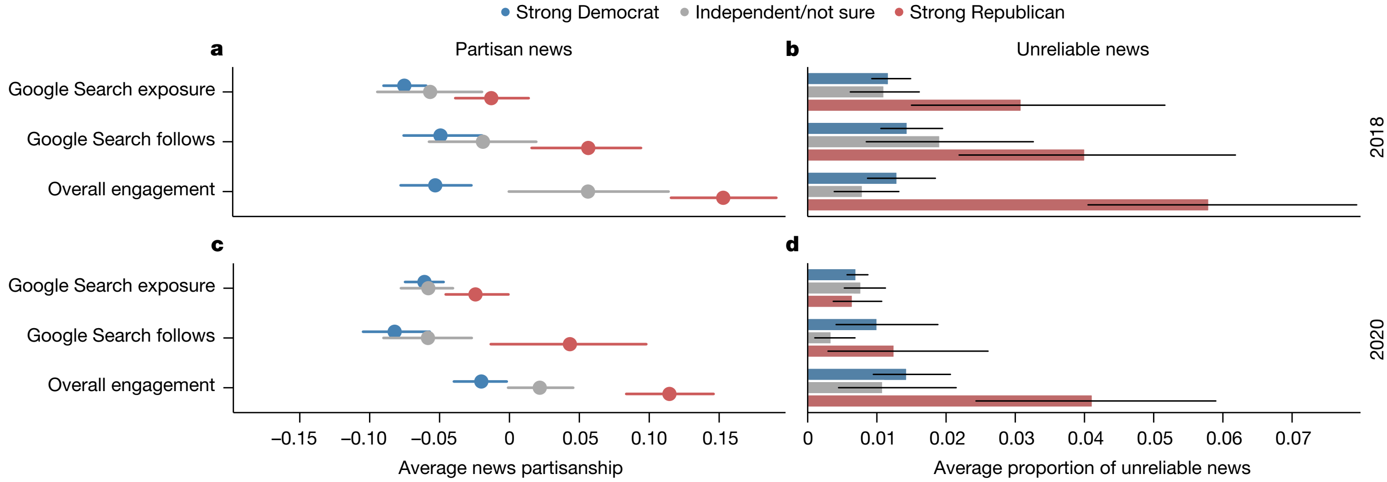


Fig. 2 | Strong partisans are exposed to similar rates of partisan and unreliable news, but asymmetrically follow and engage with such news. **a**–**d**, Average exposure, follows and overall engagement with partisan (**a**,**c**) and unreliable (**b**,**d**) news by 7-point PID clustered at the participant level in 2018 (**a**,**b**) and 2020 (**c**,**d**). The smaller differences in participants' exposure to partisan and unreliable news on Google Search aligns with previous work that found substantial homogeneity among the news domains that Google returns for controlled query sets, which tend to include mainstream national sources that have slightly left-of-zero scores in the metric we used[21,22,60]. Because a score of zero does not imply neutrality[19], left-of-zero scores do not imply a left-leaning bias (see Methods, 'Partisan news scores'). Data are presented as participant-level means grouped by 7-point PID in each subplot and all error bars indicate 95% confidence intervals (CIs). We tested for significant differences in partisan and unreliable news exposure, follows and overall engagement using the Kruskal–Wallis *H* test (Extended Data Table 2). For visual clarity, we show only a subset of partisan identity groups here (all groups are shown in Extended Data Fig. 1). For a closer look at the partisan news domains that our participants were shown or visited, we also show individual-level distributions, grouped by partisan identity and study wave, of partisan news scores in participants' exposure, follows or overall engagement (Extended Data Fig. 3).

2020) and overall engagement (12.0% in 2018, 11.9% in 2020). Gini coefficients calculated for each study wave and dataset suggest a similar pattern (Supplementary Information 1.3). These findings suggest that interactions with unreliable news are driven by a relatively small number of individuals, especially for Google Search follows in our 2020 wave and overall engagement in both waves.

Partisan and unreliable news

For each data type and study wave, we examined the relationship between partisan and unreliable news by comparing participant-level averages (Fig. 4) and using Spearman's rank correlation coefficient (ρ) to assess statistical significance (see Methods, 'Descriptive analysis'). For overall engagement, we found a significant positive correlation between partisan and unreliable news among strong Republicans in 2018 ($\rho = 0.336$; $P = 0.008$; $n = 89$) and a significant negative correlation among strong Democrats in both 2018 ($\rho = -0.380$; $P < 0.001$; $n = 115$) and 2020 ($\rho = -0.237$; $P = 0.007$; $n = 191$). By contrast, we found no significant correlations between partisan and unreliable news for strong Republicans or strong Democrats for exposure or follows in either study wave (Extended Data Table 3). These correlations suggest that, on average, partisans who choose to engage with more identity-congruent news also choose to engage with more unreliable news, but this association does not carry over into their Google Search results or follows.

Partisan news regressions

To further examine participants' partisan news outcomes, we conducted a series of multivariate regressions that controlled for extra demographic variables and theoretically motivated factors (see Methods, 'Multivariate regressions'). Across all data types—exposure, follows and overall engagement—the theoretically motivated factors we included were partisan identity and age group, and the extra demographic variables were race, gender, education and news interest. To evaluate our theoretically motivated factors, we represented age group as a four-category factor variable with 18–24 years as the reference

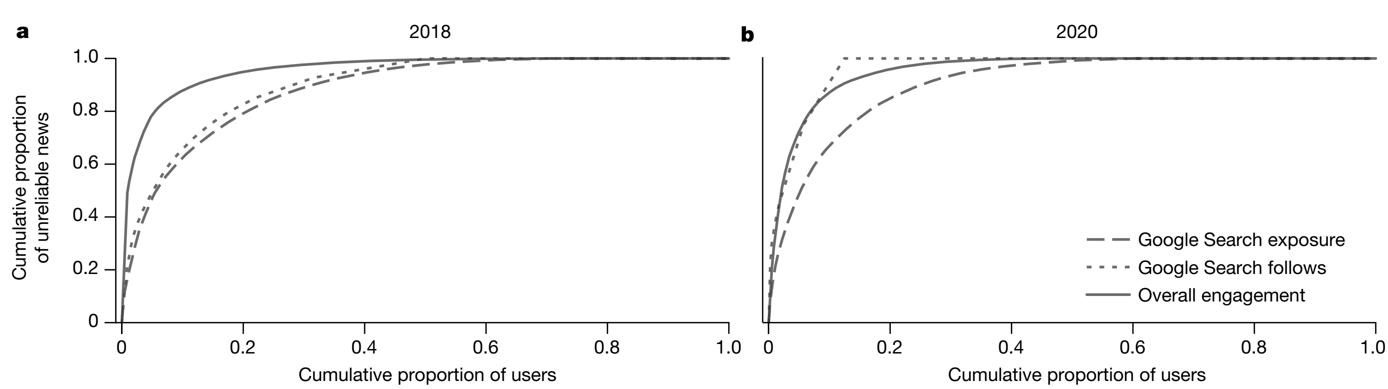


Fig. 3 | Exposure to unreliable news is less concentrated among a small number of participants than follows or overall engagement. **a**,**b**, Each line shows the cumulative proportion of participants (*x*-axis) that account for the cumulative proportion of unreliable news seen by all participants (*y*-axis) within each data type (legend) in 2018 (**a**) and 2020 (**b**).

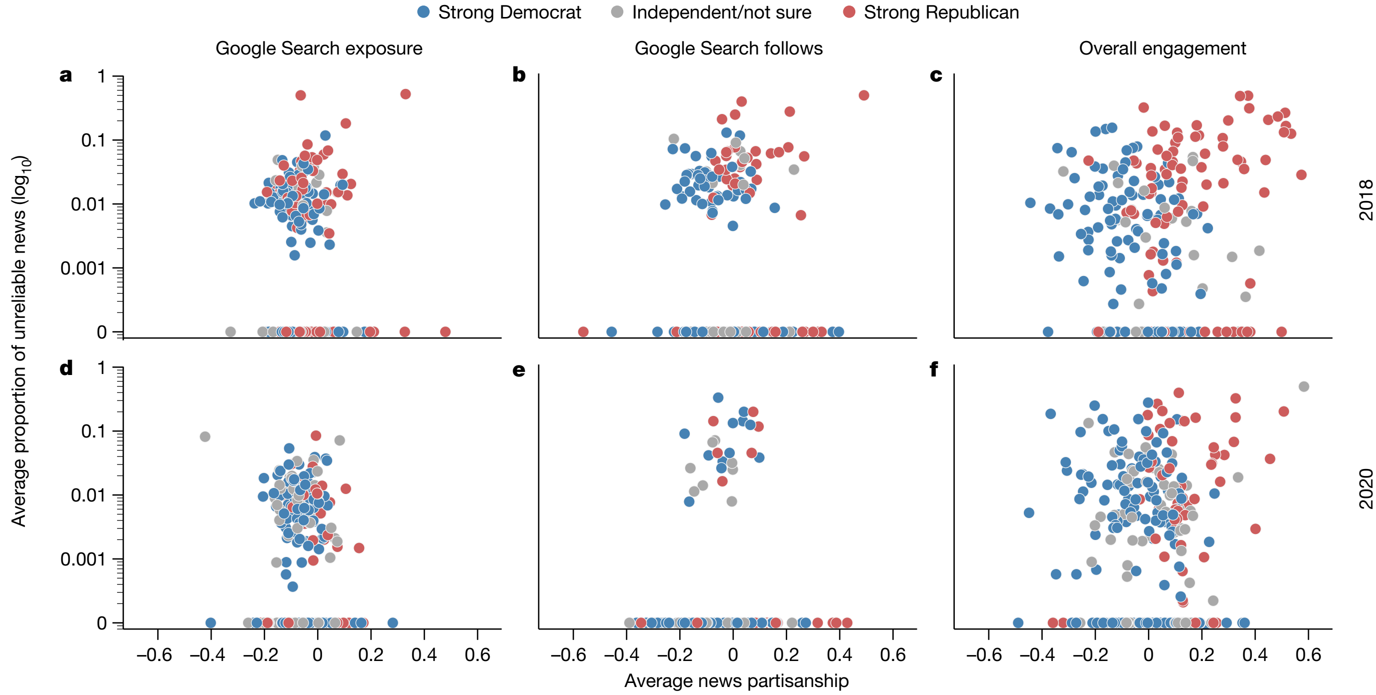


Fig. 4 | Partisans who engage with more identity-congruent news also tend to engage with more unreliable news. **a**–**f**, The relationship between partisan and unreliable news for participants' exposure on Google Search (**a**,**d**), follows from Google Search (**b**,**e**) and overall engagement (**c**,**f**) in 2018 (**a**–**c**) and 2020 (**d**–**f**). These plots highlight how the relationship between partisan and unreliable news varies across data types, and within data types when taking partisan identity into account. For visual clarity, we show only a subset of partisan identity groups here (all groups are shown in Extended Data Fig. 2). Correlations for each group and subplot are available in Extended Data Table 3.

point, and partisan identity as a seven-category factor variable (7-point PID) with 'independent' as the reference category. For exposure and follows on Google Search, we also controlled for participants' search queries by using pivoted text scaling[50] to identify their query text features in an unsupervised manner (see Methods, 'Search query analysis').

For exposure to partisan news on Google Search by partisan identity (Fig. 5a,b), we found no significant differences between strong partisans and independents in either study wave (Extended Data Tables 4 and 5). Compared with participants aged 18–24, all age groups in our 2018 wave were exposed to significantly more right-leaning news on Google Search (Extended Data Table 4), with the largest of such differences occurring for participants' aged 65+ ($b = 0.115$, (95% CI 0.064, 0.166), $P < 0.001$). In our 2020 wave (Extended Data Table 5), only the 65+ age group was exposed to significantly more right-leaning news than the 18–24 age group ($b = 0.036$ (95% CI 0.006, 0.066), $P = 0.036$). Together, these results indicate that, to the extent that there are differences in exposure to partisan news via Google Search results, they are small, inconsistent and at least partially attributable to differences in users' search queries and demographic characteristics rather than their partisan identities.

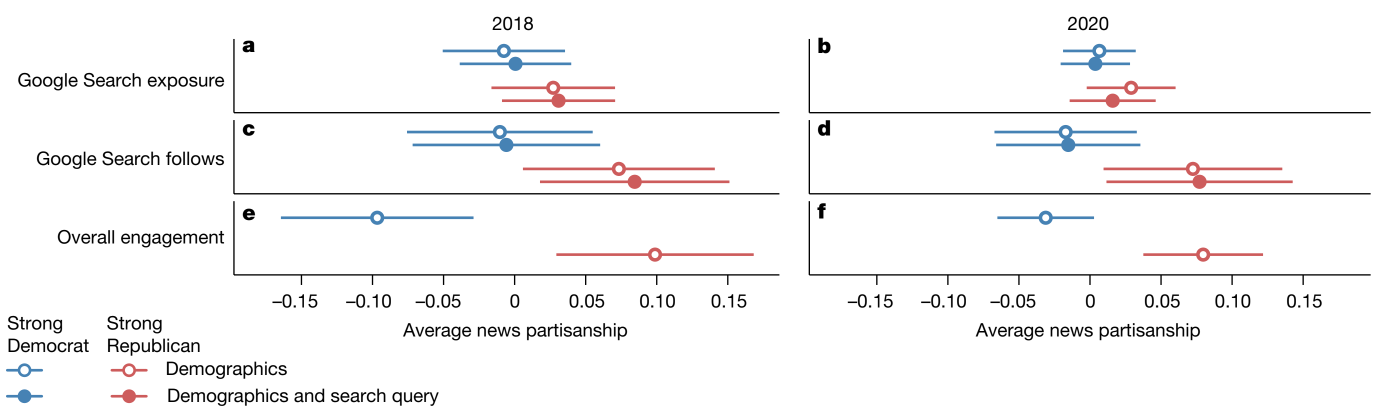


Fig. 5 | Differences in strong partisans' exposure to partisan news were not significant when controlling for demographics and search queries. **a**–**f**, Regression coefficients and 95% CIs showing the relationship between partisan identity (with independents as the reference category) and partisan news for Google Search exposure (**a**,**b**), Google Search follows (**c**,**d**) and overall engagement (**e**,**f**) in 2018 (**a**,**c**,**e**) and 2020 (**b**,**d**,**f**) study waves. Results are shown for our two primary model specifications (see Methods, 'Multivariate regressions'), which control for participants' demographics ('Demographics') or their demographics and query text features ('Demographics and search query'). For overall engagement, we examined only the demographics model because overall engagement does not directly depend on participants' search queries. Detailed results for these regressions, including estimates for all partisan identity and age groups, are available for both 2018 (Extended Data Table 4) and 2020 (Extended Data Table 5).

For participants' follows from Google Search by partisan identity (Fig. 5c,d), our regressions accounting for query text features show that strong Republicans followed significantly more right-leaning sources than independents in both 2018 ($b = 0.085$ (95% CI 0.018, 0.151), $P = 0.026$) and 2020 ($b = 0.077$ (95% CI 0.012, 0.143), $P = 0.042$). However, strong Democrats did not exhibit similar significant differences in their follows for either study wave (Extended Data Tables 4 and 5). Similar to the pattern we found for exposure, older participants followed more right-leaning news in both study waves (Extended Data Tables 4 and 5), with significant differences for participants in the 45–64 age group for 2018 ($b = 0.095$ (95% CI 0.020, 0.171), $P = 0.028$) and the 65+ age group for 2020 ($b = 0.092$ (95% CI 0.030, 0.153), $P = 0.007$). These results show that right-leaning partisans, but not left-leaning ones, are more likely to follow identity-congruent news sources from Google Search, even when accounting for the contents of their search queries.

For overall engagement with partisan news by partisan identity (Fig. 5e,f), we found that both strong Democrats and strong Republicans engaged with significantly more identity-congruent news than independents did. When accounting for all control variables, strong Democrats chose to engage with more left-leaning news than independents in 2018 ($b = -0.097$ (95% CI −0.164, −0.029), $P = 0.011$), but not in 2020 ($b = -0.031$ (95% CI −0.065, 0.003), $P = 0.145$), whereas strong Republicans chose to engage with more right-leaning news than independents in both 2018 ($b = 0.099$ (95% CI 0.029, 0.168), $P = 0.005$) and 2020 ($b = 0.080$ (95% CI 0.038, 0.122), $P < 0.001$). Age was again significantly associated with more right-leaning news (Extended Data Tables 4 and 5), and those aged 65+ had the greatest association in both 2018 ($b = 0.150$ (95% CI 0.069, 0.231), $P < 0.001$) and 2020 ($b = 0.068$ (95% CI 0.026, 0.110), $P = 0.003$). We did not include query text features in these regressions because overall engagement does not directly depend on participants' Google Search queries. In contrast to the asymmetry we observed in our participants' follows, these results indicate that participants on both the left and the right choose to engage with a significant amount of identity-congruent news sources during their overall web browsing.

In these regressions on partisan news, we also found robust associations among participants who identified as lean Republican and lean Democrat (Extended Data Tables 4 and 5). When controlling for query text features, participants in the lean Republican group were exposed to significantly more right-leaning news than independents in 2018 ($b = 0.093$ (95% CI 0.031, 0.156), $P = 0.007$). We also found significant identity-congruent associations for overall engagement with partisan news in 2020 for both lean Democrats ($b = -0.054$ (95% CI −0.093, −0.015), $P = 0.013$) and lean Republicans ($b = 0.094$ (95% CI 0.041, 0.147), $P = 0.001$). However, a relatively small number of participants identified as lean Republican in 2018 ($n = 11$) and 2020 ($n = 35$). These findings align with previous work that highlights how individuals who identify as independents, but report leaning towards one party or the other, often exhibit highly partisan attitudes and behaviours[52].

Unreliable news regressions

When accounting for both extra demographic characteristics and query text features, we found no significant differences in exposure or follows to unreliable news on Google Search by partisan identity or age group in either 2018 (Extended Data Table 6) or 2020 (Extended Data Table 7). By contrast, for overall engagement with unreliable news, we found that strong Republicans engaged with significantly more news from unreliable sources than independents did in both 2018 ($b = 1.268$ (95% CI 0.497, 2.039), $P = 0.003$) and 2020 ($b = 1.141$ (95% CI 0.529, 1.752), $P < 0.001$). No such relationship emerged for strong Democrats in either year (Extended Data Tables 6 and 7). Similar to our partisan news regressions, we also found greater overall engagement with unreliable news among participants who identified as lean Republican in 2018 ($b = 1.894$ (95% CI 0.817, 2.970), $P = 0.001$), although the number of participants in that group was again small ($n = 15$). Last, age was also associated with greater overall engagement with unreliable news in 2020 (Extended Data Table 7); compared with the 18–24 age group, those aged 45–64 ($b = 1.014$ (95% CI 0.399, 1.630), $P = 0.003$) and 65+ ($b = 1.067$ (95% CI 0.413, 1.721), $P = 0.003$) both engaged with significantly more unreliable news, but we found no significant differences by age in 2018 (Extended Data Table 6). These results show that our participants were exposed to and followed similar amounts of unreliable news on Google Search, but strong Republicans consistently chose to engage with more of it in their overall web browsing.

Discussion

The two waves of our study replicate the same finding: engagement outweighs exposure to partisan and unreliable news within Google Search, and the small differences we observe in exposure are largely explained by participants' demographic characteristics and query selection. This pattern is consistent across data collected during two distinct time periods, each with a different social, political and technological context. For concerns related to filter bubbles and echo chambers, our results highlight the role of user choice, rather than algorithmic curation, in driving such effects. These findings add to the limited number of studies examining ecological exposure[24], align with previous work on Facebook's News Feed[7] and are consistent with studies that have found engagement with unreliable or problematic content to be rare and concentrated among a small number of individuals[11,23,44].

These findings do not necessarily imply that the design of Google Search's algorithms is normatively unproblematic. In some cases, our participants were exposed to highly partisan and unreliable news on Google Search, and past work suggests that even a limited number of such exposures can have substantial negative impacts[40,53]. Moreover, the effects of such exposures may operate indirectly (for example by lowering trust in news media, rather than increasing belief in false claims) and persist over time[54,55]. However, determining the circumstances under which such content should be shown or omitted is a complex, evolving and, at times, adversarial challenge that often requires substantial social context and subject matter expertise[30,56]. As data-sharing agreements between academic institutions and industry remain limited and unstable, our approach for collecting ecologically valid exposure and engagement data directly from real users may provide a useful avenue towards independent assessments of algorithmic accountability on any platform[57].

Relatively small sample sizes present a challenge inherent to all studies that require participants to install software documenting their online behaviour[58]. Our samples were relatively small, and in some respects not representative of US population estimates, and a non-random subset of these respondents opted in to installing the browser extension (see Methods, 'Opinion surveys and participant samples'). However, given that the main findings largely replicated across both study waves, the differences between the two samples—which had different survey vendors (YouGov in 2018, PureSpectrum in 2020), different sampling procedures (oversampling strong partisans in 2018, matching national demographics in 2020), were conducted during two different election cycles (US midterm elections in 2018, US presidential election in 2020) and had slightly different data collection approaches because of changes in data accessibility (see Methods, 'Exposure and engagement datasets')—help demonstrate the robustness of our results.

One technical limitation of our study is that we relied on domain-level metrics to identify and score partisan and unreliable news sites, which limited our ability to detect differences that occur within a given domain (for example, unreliable news from *The New York Times*). We also collected data only from desktop computers, whereas mobile devices increasingly serve as an avenue to online news[59], and measured exposure only on Google Search, which accounts for most, but not all, search traffic (Supplementary Information Table 4). Similarly, we collected only the first page of search results because most searchers do not navigate beyond that point[10], but collecting results past the first page could provide further context on algorithmic curation. Further research is needed to examine the roles of other popular search engines and search on mobile devices, as well as searches conducted outside the USA, in other languages or on other socially important topics.

In the context of the broader online ecosystem, more research is needed to examine how user choice and algorithmic curation evolve and mutually shape one another over time[2]. However, research has shown that personalization on Google Search is minimal[17,19], potentially limiting the impact of such feedback loops on this platform. More research is also needed to examine how sources of influence that our datasets did not allow direct comparisons to—including messaging applications, social media, and, looking forward, AI chatbots—can affect user choice and measures of engagement. Despite these limitations, our study provides an ecologically valid look at exposure and engagement within Google Search, advances methods for collecting such data on any platform and demonstrates the importance of measuring both user choice and algorithmic curation when studying online platforms[2,7].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-023-06078-5.

1. Wagner, C. et al. Measuring algorithmically infused societies. *Nature* **595**, 197–204 (2021).
2. Rahwan, I. et al. Machine behaviour. *Nature* **568**, 477–486 (2019).
3. Sunstein, C. R. *Republic.com* (Princeton Univ. Press, 2001).
4. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. The echo chamber effect on social media. *Proc. Natl Acad. Sci.* **118**, e2023301118 (2021).
5. Lewandowsky, S., Ecker, U. K. H. & Cook, J. Beyond misinformation: understanding and coping with the 'post-truth' era. *J. Appl. Res. Mem. Cogn.* **6**, 353–369 (2017).
6. Pariser, E. *The Filter Bubble: What the Internet is Hiding From You* (Penguin, 2011).
7. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
8. Cardenal, A. S., Aguilar-Paredes, C., Galais, C. & Pérez-Montoro, M. Digital technologies and selective exposure: how choice and filter bubbles shape news media exposure. *Int. J. Press/Politics* **24**, 465–486 (2019).
9. Fletcher, R., Kalogeropoulos, A. & Nielsen, R. K. More diverse, more politically varied: how social media, search engines and aggregators shape news repertoires in the United Kingdom. *New Media & Society* https://doi.org/10.1177/14614448211027393 (2021).
10. Ekström, A. G., Niehorster, D. C. & Olsson, E. J. Self-imposed filter bubbles: selective attention and exposure in online search. *Comput. Hum. Behav. Reports* https://doi.org/10.1016/j.chbr.2022.100226 (2022).
11. Hosseinmardi, H. et al. Examining the consumption of radical content on YouTube. *Proc. Natl Acad. Sci.* **118**, e2101967118 (2021).
12. Chen, W., Pacheco, D., Yang, K.-C. & Menczer, F. Neutral bots probe political bias on social media. *Nat. Commun.* **12**, 5580 (2021).
13. Guess, A. M., Nyhan, B. & Reifler, J. Exposure to untrustworthy websites in the 2016 US election. *Nat. Hum. Behav.* **4**, 472–480 (2020).
14. Allen, J., Howland, B., Mobius, M., Rothschild, D. & Watts, D. J. Evaluating the fake news problem at the scale of the information ecosystem. *Sci. Adv.* **6**, eaay3539 (2020).
15. Muise, D. et al. Quantifying partisan news diets in Web and TV audiences. *Sci. Adv.* **8**, eabn0083 (2022).
16. Garimella, K., Smith, T., Weiss, R. & West, R. Political polarization in online news consumption. *Proc. International AAAI Conference on Web and Social Media* **15**, 152–162 (2021).
17. Hannak, A. et al. Measuring personalization of web search. In *Proc. 22nd International Conference on World Wide Web*, 527–538 https://doi.org/10.1145/2488388.2488435 (ACM, 2013).
18. Metaxa, D., Park, J. S., Landay, J. A. & Hancock, J. Search media and elections: a longitudinal investigation of political search results. *Proceedings of the ACM on Human–Computer Interaction* **3**, 129 (2019).
19. Robertson, R. E. et al. Auditing partisan audience bias within Google Search. *Proc. ACM on Human–Computer Interaction* **2**, 148 (2018).
20. Trielli, D. & Diakopoulos, N. Partisan search behavior and Google results in the 2018 U.S. midterm elections. *Inf. Commun. Soc.* **25**, 145–161 (2020).
21. Fischer, S., Jaidka, K. & Lelkes, Y. Auditing local news presence on Google News. *Nat. Hum. Behav.* **4**, 1236–1244 (2020).
22. Kawakami, A., Umarova, K. & Mustafaraj, E. The media coverage of the 2020 US presidential election candidates through the lens of Google's top stories. In *Proc. International AAAI Conference on Web and Social Media* **14**, 868–877 (AAAI, 2020).
23. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. Fake news on Twitter during the 2016 U.S. presidential election. *Science* **363**, 374–378 (2019).
24. Huszár, F. et al. Algorithmic amplification of politics on Twitter. *Proc. Natl Acad. Sci.* **119**, e2025334119 (2022).
25. Lazer, D. M. J. et al. The science of fake news. *Science* **359**, 1094–1096 (2018).
26. Watts, D. J., Rothschild, D. M. & Mobius, M. Measuring the news and its impact on democracy. *Proc. Natl Acad. Sci.* **118**, e1912443118 (2021).
27. Gillespie, T. The politics of 'platforms'. *New Media & Society* **12**, 347–364 (2010).
28. Taber, C. S. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. *Am. J. Political Sci.* **50**, 755–769 (2006).
29. Iyengar, S. & Hahn, K. S. Red media, blue media: evidence of ideological selectivity in media use. *J. Commun.* **59**, 19–39 (2009).
30. Introna, L. D. & Nissenbaum, H. Shaping the web: why the politics of search engines matters. *The Information Society* **16**, 169–185 (2000).
31. Lawrence, S. & Giles, C. L. Accessibility of information on the web. *Nature* **400**, 107–107 (1999).
32. Metaxas, P. T. & DeStefano, J. Web spam, propaganda and trust. In *Proc. 2005 World Wide Web Conference* (2005).
33. Vaidhyanathan, S. *The Googlization of Everything (And Why We Should Worry)* (Univ. of California Press, 2011).
34. Allcott, H. & Gentzkow, M. Social media and fake news in the 2016 election. *J. Econ. Perspect.* **31**, 211–236 (2017).
35. Newman, N., Fletcher, R., Kalogeropoulos, A. & Neilsen, R. K. *Digital News Report 2019* (Reuters Institute for the Study of Journalism, 2019).
36. Mitchell, A., Gottfried, J., Shearer, E. & Lu, K. *How Americans Encounter, Recall and Act Upon Digital News* (Pew Research Center, 2017).
37. Edelman. *Edelman Trust Barometer 2021* (2021).
38. Golebiewski, M. & Boyd, D. *Data Voids: Where Missing Data can Easily be Exploited* (Data & Society, 2019).
39. Tripodi, F. *Searching for Alternative Facts: Analyzing Scriptural Inference in Conservative News Practices* (Data & Society, 2018).
40. Epstein, R. & Robertson, R. E. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proc. Natl Acad. Sci.* **112**, E4512–E4521 (2015).
41. Epstein, R., Robertson, R. E., Lazer, D. & Wilson, C. Suppressing the search engine manipulation effect (SEME). *Proc. ACM on Human–Computer Interaction* **1**, 42 (2017).
42. Wojcieszak, M., Menchen-Trevino, E., Goncalves, J. F. F. & Weeks, B. Avenues to news and diverse news exposure online: comparing direct navigation, social media, news aggregators, search queries, and article hyperlinks. *Int. J. Press/Politics* **27**, 860–886 (2022); https://doi.org/10.1177/19401612211009160.
43. Peterson, E., Goel, S. & Iyengar, S. Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign. *Political Sci. Res. Methods* **9**, 242–258 (2021).
44. Guess, A. M., Nagler, J. & Tucker, J. Less than you think: prevalence and predictors of fake news dissemination on Facebook. *Sci. Adv.* **5**, eaau4586 (2019).
45. Yin, L. Local news dataset. *Zenodo* https://doi.org/10.5281/ZENODO.1345145 (2018).
46. Rating Process and Criteria. *NewsGuard* https://www.newsguardtech.com/ratings/rating-process-criteria/ (2021).
47. Mustafaraj, E., Lurie, E. & Devine, C. The case for voter-centered audits of search engines during political elections. In *Proc. 2020 Conference on Fairness, Accountability, and Transparency* 559–569 (ACM, 2020).
48. van Hoof, M., Meppelink, C. S., Moeller, J. & Trilling, D. Searching differently? How political attitudes impact search queries about political issues. *New Media & Society* https://doi.org/10.1177/14614448221104405 (2022).
49. Zuckerman, E. Why study media ecosystems? *Inf. Commun. Soc.* **24**, 1495–1513 (2021).
50. Hobbs, W. R. Text scaling for open-ended survey responses and social media posts. *SSRN* https://doi.org/10.2139/ssrn.3044864 (2019).
51. Flaxman, S., Goel, S. & Rao, J. M. Filter bubbles, echo chambers, and online news consumption. *Public Opin. Q.* **80**, 298–320 (2016).
52. Klar, S. & Krupnikov, Y. *Independent Politics: How American Disdain for Parties Leads to Political Inaction* (Cambridge Univ. Press, 2016).
53. Noble, S. U. *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York Univ. Press, 2018).
54. Guess, A. M., Barberá, P., Munzert, S. & Yang, J. The consequences of online partisan media. *Proc. Natl Acad. Sci.* **118**, e2013464118 (2021).
55. Kobayashi, T., Taka, F. & Suzuki, T. Can 'Googling' correct misbelief? Cognitive and affective consequences of online search. *PLoS ONE* **16**, e0256575 (2021).
56. Gillespie, T. Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. *Inf. Commun. Soc.* **20**, 63–80 (2017).
57. Diakopoulos, N. Algorithmic accountability: journalistic investigation of computational power structures. *Digit. Journal.* **3**, 398–415 (2015).
58. Parry, D. A. et al. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nat. Hum. Behav.* **5**, 1535–1547 (2021).
59. Walker, M. Americans favor mobile devices over desktops and laptops for getting news. *Pew Research Center* https://www.pewresearch.org/short-reads/2019/11/19/americans-favor-mobile-devices-over-desktops-and-laptops-for-getting-news/ (2019).
60. Trielli, D. & Diakopoulos, N. Search as news curator: the role of Google in shaping attention to news information. In *Proc. 2019 CHI Conference on Human Factors in Computing Systems* https://doi.org/10.1145/3290605.3300683 (ACM Press, 2019).

Methods

Opinion surveys and participant samples

The self-reported data used in this study were based on two multi-wave public opinion surveys. In each of them, respondents were asked to install a browser extension that would monitor their Google Search results and various aspects of their online activity. The protocol and informed consent language we used while recruiting subjects was transparent about what the extension collected, and both studies were approved by the Institutional Review Board (IRB) at Northeastern University (no. 18-10-03 for 2018, and no. 20-03-04 for 2020).

The first survey was fielded between 18 and 24 October 2018, and participants were recruited through YouGov with an oversample of strong partisans. The second was fielded between June 2020 and January 2021, with participants recruited by PureSpectrum as part of the 'COVID States Project' (https://covidstates.org). This second survey used quota sampling based on state-by-state benchmarks for race, gender and age, but a non-random subset of these respondents opted in to installing our browser extension.

In both surveys, participants self-reported a 7-point PID (strong Democrat, not very strong Democrat, lean Democrat, independent, lean Republican, not very strong Republican, strong Republican); their age, which we binned (18–24, 25–44, 45–64, 65+); and several other demographic variables (race, gender, education). We use a seven-point partisan scale, rather than a three-point scale (Democrat, independent, Republican) for two reasons. First, binning all Republicans and Democrats into a single category would obscure our intentional oversampling of strong partisans in 2018. Second, recent work in political science has found that independents who lean towards a political party tend to behave in a more partisan manner than weak partisan identifiers[52], further complicating the use of a three-point scale.

Owing to budget constraints around sample size, we anticipated a relatively small sample size in 2018, and intentionally oversampled on strong partisans. As a result, approximately 36% of our participants identified as strong Democrats and 27% identified as strong Republicans (Supplementary Information Table 1). Ensuring that we had a significant number of such people in our sample was key to our investigation because engagement with partisan and unreliable news has been shown to be uncommon and concentrated among a relatively small number of predominantly partisan individuals[13,23,44]. In 2020, respondents were selected by means of quota sampling based on population benchmarks in US states for several categories (race, gender and age); 27% identified as strong Democrats and 13% identified as strong Republicans (Supplementary Information Table 2). We did not conduct a power analysis before either study wave because of a combination of budget constraints, uncertainty around how many survey participants would opt in to installing our browser extension (Supplementary Information 1.1) and uncertainty around potential effect sizes. However, the consistency of the main results across both study waves suggests that our sample sizes were sufficiently powered.

Exposure and engagement definitions

We broadly define exposure as a measure of the content appearing on a user's screen, including links, text and media. In this study, we specifically measure exposure to links that appeared on a participant's screen after they conducted a Google search. Our approach and definition for exposure are related to concepts including 'involuntary exposure'[8], 'passive exposure'[14], 'distributed access'[9], 'incidental exposure'[61], 'linger impressions'[24] and other uses of 'exposure'[62,63], but most closely align with Bakshy et al.[7], who defined an 'exposure' as a case "in which a link to the content appears on screen in an individual's [Facebook] News Feed"[7]. Our approach for capturing exposure is also analogous to that used by the Screenome project, which preserves what users saw through the "collection of high-density sequences of screenshots"[64]. However, instead of periodically collecting image screenshots, we took HTML snapshots when participants visited a given website, which has the advantage of enabling programmatic extraction of URLs that appeared on the screen, but does not guarantee they were seen.

We broadly define engagement as a measure of the actions taken by a user in their web browser, including URLs clicked, typed or copied–pasted, and browser tab switches. Overall engagement therefore provides an aggregate record of users' online choices, regardless of how those choices were arrived at (including the influence of other online platforms), which we measured using several digital trace datasets that we collected through existing and custom data collection tools (see 'Exposure and engagement datasets'). Our use of engagement aligns with terms from previous work including "visits", "clicks" and "consumption"[7,14,63], "direct access"[9], and "choice" or "voluntary exposure"[8]. These terms share an emphasis on user choice, whereas 'engagement' expands the scope of behaviours considered and provides a high-level term for digital trace logs with varying granularity but common measurement aims.

Using the above definitions, we broadly define a follow as an instance in which a person is exposed to search results, a news feed or recommendations during a visit to an online platform, and subsequently visits a URL within a time threshold and domain matching condition when possible (see 'Exposure and engagement datasets'). Follows therefore represent the subset of overall engagement that can be attributed to on-platform choices. Although this term overlaps with social media terms for 'following' an account or channel, which can be ambiguous and vary by platform (for example, channel 'subscriptions' on YouTube), it also accurately describes the process of following a hyperlink from one web page to another. Our approach is in line with previous work on identifying follows, which have previously been called "referrals"[14], and their preceding platform visits "referrers"[13]. For example, a visit to bbc.com might be attributed to Facebook if a visit to Facebook preceded that visit, and occurred within a short (for example 1 min) time threshold[9,14,51,65]. Another approach involves examining news article visits and checking if a platform visit occurred within the past three URLs visited and previous 30 s (ref. 13). We use follows instead of referrals here because the term 'referral' collides with the HTTP 'referer' header field (which we did not collect), which may contain information about the previously visited website, but has evolved over time, is omitted under certain circumstances and can be modified by both the user and the 'referring' platform.

Exposure and engagement datasets

We built custom browser extensions in 2018 and 2020 to collect several types of digital trace data that would enable us to compare the URLs participants were shown on Google Search (exposure), the results they followed after that Google Search exposure (follows) and the URLs they engaged with on the web in general (overall engagement). In terms of exposure, the extension passively, by monitoring the participant's web browsing activity, collected snapshots of their Google Search results. For overall engagement, the extension actively, through automated periodic requests, collected each participant's Browser History. To measure follows in 2018, the extension also actively collected their Google History, which consists of clicks on Google Search and related Google services, but did not collect the same data in 2020 because of changes in how that website functioned. To measure follows in 2020, we introduced a passive measure of engagement that tracked the participant's active browser tab, which we call 'Tab Activity'.

In both waves, we filtered out participants who did not meet certain web activity levels. Specifically, we filtered out participants whose participation window—the duration between their first and last observed behaviour—was less than 10 days, which suggests that they installed and then quickly uninstalled the extension. For the 2018 datasets, this led to filtering out the Browser History of 7 participants, the Google History for 5 participants and Google Search results for 30 participants. Using the same rules for our 2020 datasets, we filtered out Browser

History for 25 participants, Tab Activity for 197 participants and Google Search results for 150 participants. These filtering decisions did not substantially remove more participants from one 7-point PID group than another. We report overall counts, including the number of searches conducted and URLs clicked, in Extended Data Table 1.

In accordance with IRB guidelines, full consent was obtained, and participants were compensated for installing the extension. Participants were informed that they could uninstall the extension at any time, and at the end of the data collection period, the extension automatically uninstalled itself. The extension did not collect any activity that occurred in incognito (private) browser windows. Further details on our IRB study procedures are available in Supplementary Information 1.5.

Google Search results. For each search the participants conducted while our browser extension was installed, we saved an HTML snapshot of the corresponding SERP that they were exposed to on Google Search. We collected only the first SERP because most search traffic goes to top-ranked results, and most users do not go past the first page of results[66]. The extension did not collect snapshots of participants' Google Searches in incognito (private) browser windows. To identify the URLs participants were exposed to, we parsed each HTML snapshot and extracted detailed information from each result, including a URL (if present) and a classification of the result format (for example, news, Twitter or knowledge results). The SERP parser we used is available at: https://github.com/gitronald/WebSearcher.

Google History. For the Google History dataset, we collected participants' Google account activity, which provided us with a log of their clicks on Google SERPs, as well as their activity on Google News and other Google products. We collected these data by identifying an unofficial API end point and then incrementally collecting data through requests periodically sent by the browser extension. These requests occurred every 2 weeks throughout the study. To ensure that we could access that page, we asked participants to remain logged in for the duration of the study and sent them reminders to log in if they were not when the extension periodically attempted to collect the latest data. Although the interface has changed since 2018, data similar to the types we collected can be seen on Google's My Activity page (https://myactivity.google.com) while logged in to a Google account. Similar to past work[13,65], we identified and removed consecutive visits to the same web page that occurred within 1 s, keeping only the first instance. Consecutive visits to the same web page are often present in such data, and can occur for a variety of reasons, such as refreshing a stalled page or website-specific idiosyncrasies in page loading. Using the 253 days of Google History data that we were able to collect (Extended Data Table 1), we identified follows by first identifying each search query logged in a participant's Google History and then counting the subsequent website visit logged in their Google History as a follow, conditional on if it occurred within 60 s of the search[9,14,51,65].

Tab Activity. In the 2020 extension, we designed Tab Activity to log changes in the active browser tab. This overcomes limitations with current approaches to collecting engagement data. For example, it is often unclear exactly how data from proprietary sources are obtained, logged or cleaned. By contrast, Browser History, which is maintained by Chrome and Firefox, has public documentation that suggests it provides a log of the first time a web page loads in the browser, but does not account for changes in the active browser tab. By passively monitoring the active browser tab, Tab Activity provides a more direct and detailed measure of user attention than other engagement measures, such as Browser History, by accounting for the common practice of tabbed browsing, which can affect measurements of URL visits in terms of both time and volume. This enabled us to more directly link exposures to engagements, which we used to identify follows as URL visits when (1) they occurred immediately after and within 60 s of an exposure and (2) the domain they followed matched any of the search results shown to them during that exposure. When adding this second condition, we were able to match 33% of all follows, and 62% of news follows, to a corresponding search result. We removed duplicates from Tab Activity using the same method as we did in Google History, removing sequential duplicates that occurred within a 1-s interval of each other.

Browser History. We collected Browser History by accessing an API that is built in to Chrome and Firefox. In both data collection waves, the extension collected data every 2 weeks for the duration of the study. An important difference between the Browser History provided by Chrome and Firefox is that Chrome allows you to access only the past 3 months of Browser History, whereas Firefox allows you to continue saving history regardless of time passed, but enforces a maximum page limit. In 2018, we collected a pre-aggregated version of Browser History (HistoryItems), whereas in 2020 we collected higher-resolution data that captured each website visit (VisitItems). Further details on the Browser History API are available in Mozilla's documentation (https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/API/history). For all analyses, we combined visits and typed counts into a total count. Owing to the API-level aggregation of the 2018 Browser History data, we did not apply duplicate correction to it. However, for the 2020 data we identified and removed consecutive visits to the same web page that occurred within 1 s, as we did with Google History and Tab Activity.

Web domain scores and classifications

We classified and scored online news at the domain level. That is, for each URL in our dataset, we extracted the second-level domain name (for example, https://www.cnn.com → cnn.com). This enabled us to merge our data with several datasets that contain domain-level scores and classifications for news, partisanship and reliability.

News classifications. To identify visits to news domains, we compiled four datasets containing classifications of web domains as 'news'. After cleaning, the union of these datasets gave us 11,902 unique domains coded as news. These datasets include:
(1) 488 domains identified as 'hard news' by Bakshy et al.[7]
(2) 1,250 domains manually identified as news by Grinberg et al.[23]
(3) 6,288 domains aggregated from local news listings by Yin[45]
(4) 6,117 domains identified as news by NewsGuard[46].

Bakshy et al.[7] used the URLs and associated text snippets shared by Facebook users who self-reported a partisan identity to classify 'hard news' domains, and publicly released the 500 most shared domains[7]. As in previous work using these classifications[63], we exclude their classifications of five platforms (youtube.com, m.youtube.com, amazon.com, twitter.com and vimeo.com) as news. We also excluded a satire site (theonion.com) and Wikipedia as news sites and merged five entries that had a duplicate including a 'www.' prefix, which typically direct to the same home page, leaving us with 488 domains. The list used by Grinberg et al.[23] was manually curated and contains a list of 1,250 domains coded as news on the basis of editorial practices[23].

Yin[45] provides a list of 6,288 domains associated with state newspapers, TV stations and magazines, aggregated from several sources, including the United States Newspaper Listing[45]. We removed a string value ('Alaskan Broadcast Television') and the coding of myspace.com (a social media platform) as news domains from the current version of this dataset. NewsGuard is an independent organization that "employs a team of trained journalists and experienced editors to review and rate news and information websites based on nine journalistic criteria." To maintain consistency with the other lists, we classified all domains covered by NewsGuard as news except those labelled as a satire site ("not a real news website") or a platform ("primarily hosts user-generated content that it does not vet"), as neither receives news ratings based on the nine journalistic criteria[46].

We measured the proportion of news present in Google's SERPs in two ways. At the URL level, we used the mean proportion of links on a SERP that led to news domains. At the SERP level, we used the mean proportion of SERPs containing at least one news domain. Counts for both measures are available in Extended Data Table 1, and participant-level averages are available in Supplementary Information Tables 5 and 6.

Partisan news scores. To quantify participants' exposure and engagement with partisan news, we used the partisan audience bias scores developed in previous work[19]. These scores were made using the domain-sharing patterns of a large virtual panel of Twitter users who had been linked to US voter registration records. More specifically, each domain was scored by comparing the relative proportion of registered Democrats and Republicans who shared it on Twitter. The resulting scores range from −1 (shared only by Democrats) to 1 (shared only by Republicans). On this scale, a domain score of 0 does not mean neutral or unbiased, only that an equal proportion of Democrats and Republicans in the virtual panel shared it.

These scores have been used in several recent examinations of engagement with partisan news[15,16,42], provide coverage for more domains overall (n = 19,022) and news domains specifically (n = 2,584, covering 21.7% of the 11,902 domains we classified as news) than other domain-level partisanship metrics and are strongly correlated (r = 96***) with the most widely used alternative[7]. However, similar to other domain-level metrics[7], these scores are limited in terms of the context of the shares they are based on: a user may share a news article from a domain because they are denouncing it, not because they support it.

We aggregated partisanship scores to the participant level in slightly different ways for our exposure and engagement data. For exposure to Google Search results, we calculated the partisanship of each SERP using a weighted average that takes ranking into account[19] and places more weight to the partisan audience scores of domains appearing near the top of the search rankings. This helps to account for the extra attention received by highly ranked search results[10,66–68], which is partly due to position effects that are well established in the psychological sciences[69].

Weighted average bias B^{w} is calculated by finding the bias B for a given query q, using the domain score s of each item i on the page until reaching maximun rank r: $B(q, r) = \sum_{i=1}^{r} s_i / r$, and then taking the normalized sum of this over all ranks: $B^{\mathrm{w}}(q, r) = \sum_{i=1}^{r} B(q, i)/r$. We then used those rank-weighted averages to calculate the average news partisanship of each participant's search results. For our engagement data, we calculated a participant-level average by taking the mean score of all the news domains they visited during the study. To handle the aggregated Browser History data we collected in 2018, we multiplied each domain's score by the number of visits to that domain, and then divided the sum by the total number of news visits. We report group differences throughout the paper using these participant-level averages, and provide individual-level partisan news distributions for each dataset and study wave in Extended Data Fig. 3.

Unreliable news classifications. We classified 2,962 web domains as unreliable if they appeared in either of two carefully constructed lists[23,46]. First, NewsGuard (introduced in the News classifications section) tracks more than 6,000 news websites for information quality and assigns each a score from 0 (unreliable) to 100 (reliable)[46]. We used the threshold defined by NewsGuard to classify news domains, and labelled each of the 2,534 domains with a score of less than 60 as unreliable. Second, Grinberg et al.[23] investigated fake news sharing on Twitter and manually assembled their list of unreliable news domains[23]. These classifications were made by examining fact checkers' evaluations of stories produced by various domains, and defined fake news as content that has the form of standard media, but not the intent or processes to produce accurate content. The colours they used to code domains include black if they contained "almost exclusively fabricated stories", red if they "spread falsehoods that clearly reflected a flawed editorial process" and orange in "cases where annotators were less certain that the falsehood stemmed from a systematically flawed process". As in the original paper, we consider all 490 domains coded as black, red or orange to be unreliable news domains.

Descriptive analysis

To assess overall search engine use, we identified URLs leading to popular web search engines by filtering for known domains (for example, google.com) and a URL path indicating a page of search results (for examplek google.com/search), which excludes visits to each search engine's home page, as in previous work[14]. We found that Google handled a substantial majority of our participants' web searches in both 2018 (74.2%) and 2020 (68.6%), and Bing handled the second most (21.7% in 2018; 30.5% in 2020). This majority use of Google Search is in line with industry estimates of Google's desktop market share[70,71], and provides support for our focus on this search engine. We provide further details on overall search engine use in Supplementary Information Table 4.

We report overall data counts for each study wave, including the total number of URL exposures, follows and engagements, in Extended Data Table 1. To provide further context on participants' web behaviour, we also report the average number and percentage of participants' exposures, follows and engagements that involved news domains in Supplementary Information Tables 5 and 6. For the average participant in either study wave, we found a substantially greater proportion of news in the Google Search results that they were exposed to (14.3% for 2018, 14.7% for 2020), than we did in amount of news they chose to follow (8.1% for 2018, 8.9% for 2020) or engage with overall (7.1% for 2018, 3.2% for 2020).

For the results presented in Fig. 2, we tested for statistically significant differences in partisan and unreliable news by partisan identity and age group using the Kruskal–Wallis H test. This test is a one-sided non-parametric test of differences among three or more groups where P is the survival function of the X^2 distribution evaluated at H. We used this non-parametric test because of the heterogeneity we observed in participant-level averages of news exposure, follows and engagement (Supplementary Information Tables 5 and 6), and present results for each Kruskal–Wallis test in Extended Data Table 2. We did not adjust P values for multiple comparisons in these tests because we did not conduct pairwise tests between groups (for example between strong Democrats and independents). Instead, we limited this analysis to assessing whether any differences occurred among the two factors motivated by previous work (partisan identity and age), and provide a more detailed examination of how each partisan group compared with independents in our regression analysis (Methods, 'Multivariate regressions'). We assessed statistical significance with $\alpha = 0.05$ throughout.

We evaluated the relationship between partisan and unreliable news in each dataset and study wave (Fig. 4) using another non-parametric test: Spearman's rank correlation coefficient (ρ). To calculate P values for this test, we used a two-sided t-test with $n - 2$ degrees of freedom. The only significant overall correlation we observed, in any dataset or study wave, was a negative correlation between partisan and unreliable news for exposure in 2020 (n = 453, $\rho = -0.162$, $P < 0.001$). However, owing to the visual differences we observed among partisan identity groups in Fig. 4, which suggest a Simpson's paradox, we also calculated the correlation between partisan and unreliable news for each partisan identity group in each dataset and study wave (Extended Data Table 3). After adjusting these P values for multiple comparisons across 7-point PID using the Holm–Bonferroni method, the only significant correlations were in participants' overall engagement, and were negative among strong Democrats and positive among strong Republicans (Extended Data Table 3).

Search query analysis

To quantify the content of participants' search queries, we used pivoted text scaling[50]. This method is a form of principal component analysis that is performed on a truncated word co-occurrence matrix (participants' queries) and used to identifying orthogonal latent dimensions that explain decreasing shares of variation in the co-occurrence of commonly used words. Each word can then be given a score along each dimension, and each document (a query) can be scored with respect to each dimension on the basis of its average word scores. Using those scores, we calculated each participant's average document score along each of the first nine dimensions derived from their respective corpora, and included these query text features in our regressions on participants' Google Search exposure and follows.

This measure is well suited for the purposes of our study for several reasons. First, as it is unsupervised, it does not require the use of external dictionaries of partisan speech developed in other contexts, such as politicians' speech, which may not be applicable to search queries. Second, it does not rely on extra participant-level information (such as partisan identity) that could introduce collinearity into subsequent regressions that use our quantitative representations of participants' query text. Finally, the method was designed specifically for short documents—such as social media posts or, in our case, search queries—for which other unsupervised methods such as topic models are less efficient.

Multivariate regressions

We ran a series of regressions to estimate the associations between several theoretically motivated factors and exposure to, follows to and engagement with partisan or unreliable news.

Outcomes and model selection. For partisan news, our outcome is the average partisan audience score described above, and we estimate this outcome using ordinary least squares. For unreliable news, our outcome is the count either of URLs from unreliable sources or of SERPs that contain at least one unreliable source. As these count outcomes are overdispersed, in that their variances are greater than their means, we estimate them using negative binomial regressions.

Independent variables. Drawing on previous research[13,23,44] and our descriptive results above, our primary independent variables were age and partisan identity. We also considered five extra adjustment variables for each participant: (1) race (white/non-white), (2) gender (male/female), (3) education (college/non-college), (4) news interest (high/low) and (5) query text features. Although not the focus of our study, there are limitations to using binary variables for race[72] and gender[73]. We coded high news interest on the basis of participants in 2018 self-reporting that they follow news and current events "most of the time" and participants in 2020 self-reporting that they were "very" or "extremely" interested in US politics and government. Query text features were calculated using pivoted text scaling[50] (Methods, 'Search query analysis').

We iteratively built our models around partisan identity and age because previous research identified them as being associated with exposure and engagement with partisan and unreliable news[13,23,44,63]. Therefore, we used four specifications of our independent variables in each model: (1) partisan identity, (2) age group, (3) age group and partisan identity and (4) age group, partisan identity and demographic controls. We represented age group as a four-category factor variable with 18–24 as the reference point, and partisan identity as a seven-category factor variable (7-point PID) with 'independent' as the reference category. For outcomes that depend on search results (Google Search exposure and follows), we also included a fifth model specification: (5) age group, partisan identity, demographic controls and query text features. We did not include query text features in our regressions on overall engagement outcomes because those outcomes do not directly depend on search queries.

Regression results. Regression results for each outcome—partisan and unreliable news in Google Search exposure, Google Search follows and overall engagement—are available in Extended Data Tables 4–7 and Supplementary Information Tables 7–18.

For partisan news outcomes among the 2018 sample, results from our two primary model specifications, models 4 (demographics) and 5 (demographics + query text features), are available in Extended Data Table 4. Results from all partisan news model specifications are available in the Supplementary Information for Google Search exposure (Supplementary Information Table 7), Google Search follows (Supplementary Information Table 8) and overall engagement (Supplementary Information Table 9). Similarly, for partisan news outcomes in 2020, the main results are in Extended Data Table 5, and results from all model specifications are available in the Supplementary Information for Google Search exposure (Supplementary Information Table 10), Google Search follows (Supplementary Information Table 11) and overall engagement (Supplementary Information Table 12).

For unreliable news outcomes, our main results from 2018 are available in Extended Data Table 6. All unreliable news model specifications for 2018 are available in the Supplementary Information for Google Search exposure (Supplementary Information Table 13), Google Search follows (Supplementary Information Table 14) and overall engagement (Supplementary Information Table 15). Last, for unreliable news outcomes among the 2020 sample, the main results are available in Extended Data Table 7, and full model specifications are available in the Supplementary Information for Google Search exposure (Supplementary Information Table 16), Google Search follows (Supplementary Information Table 17) and overall engagement (Supplementary Information Table 18).

All tests for statistical significance in our regressions are two-sided *t*-tests. We provide *t*-statistics in Extended Data Tables 4–7. When adjusting for multiple comparisons, we did not correct across model specifications, because different specifications within the same model are not independent, nor across data types (exposure/follows/engagement), because they are also not fully independent of one another. Because we conducted the 2020 study wave as a replication, which is another way of guarding against false discovery, we also did not correct across study waves. For outcomes (partisan and unreliable news), we have theoretical reasons to expect age to matter for unreliable news, but not for partisan news. However, to test the robustness of our results, we adjusted P values across outcomes for our primary models (Extended Data Tables 4–7) using the Holm–Bonferroni method. These adjustments did not change the significance of any unreliable news outcomes, but five partisan news outcomes were not robust to the adjustments, including differences in overall engagement for participants who identified as 25–44 or lean Democrat in 2018 (Extended Data Table 4), differences in follows for participants who identified as 65+ or lean Republican in 2018 (Extended Data Table 4) and differences in exposure for participants who identified as not very strong Republicans in 2020 (Extended Data Table 5). Results presented in Supplementary Information Tables 7–18 provide further details for each model specification we examined, including degrees of freedom and R^2 values, and are not adjusted for multiple comparisons.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Owing to privacy concerns and IRB limitations, visit-level data will not be released, but aggregated data are available at https://doi.

org/10.7910/DVN/WANAX3. The domain scores and classifications we used are available at https://github.com/gitronald/domains, but the NewsGuard classifications are not included because of their proprietary nature.

Code availability

The data for this study were collected using custom browser extensions written in JavaScript and using the WebExtension framework for cross-browser compatibility. The source code for the extensions we used in 2018 and 2020 is available at https://github.com/gitronald/webusage, and a replication package for our results is available at https://github.com/gitronald/google-exposure-engagement. The parser we used to extract the URLs our participants were exposed to while searching is available at https://github.com/gitronald/WebSearcher. Analyses were performed with Python v.3.10.4, pandas v.1.4.3, scipy v.1.8.1, Spark v.3.1 and R v.4.1.

61. Nanz, A. & Matthes, J. Democratic consequences of incidental exposure to political information: a meta-analysis. *J. Commun.* **72**, 345–373 (2022).
62. Bechmann, A. & Nielbo, K. L. Are we exposed to the same 'news' in the news feed? An empirical analysis of filter bubbles as information similarity for Danish Facebook users. *Digit. Journal.* **6**, 990–1002 (2018).
63. Guess, A. M. (Almost) everything in moderation: new evidence on Americans' online media diets. *Am. J. Political Sci.* **65**, 1007–1022 (2021).
64. Reeves, B. et al. Screenomics: a framework to capture and analyze personal life experiences and the ways that technology shapes them. *Hum. Comput. Interact.* **36**, 150–201 (2021).
65. Stier, S., Mangold, F., Scharkow, M. & Breuer, J. Post post-broadcast democracy? News exposure in the age of online intermediaries. *Am. Political Sci. Rev.* **116**, 768–774 (2021).
66. Pan, B. et al. In Google we trust: users' decisions on rank, position, and relevance. *J. Comput.-Mediat. Commun.* **12**, 801–823 (2007).
67. Kulshrestha, J. et al. Search bias quantification: investigating political bias in social media and web search. *Inf. Retr. J.* **22**, 188–227 (2019).
68. Joachims, T. et al. Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search. *ACM Trans. Inf. Syst.* **25**, 7–es (2007).
69. Ebbinghaus, H. *Memory: A Contribution to Experimental Psychology* https://doi.org/10.1037/10011-000 (Teachers College Press, 1913).
70. Fishkin, R. New jumpshot 2018 data: where searches happen on the web (Google, Amazon, Facebook, & beyond). *SparkToro* https://sparktoro.com/blog/new-jumpshot-2018-data-where-searches-happen-on-the-web-google-amazon-facebook-beyond/ (2018).
71. Desktop search engine market share United States of America. *StatCounter Global Stats* https://gs.statcounter.com/search-engine-market-share/desktop/united-states-of-america/2020 (2020).
72. Brown, N. E. Political participation of women of color: an intersectional analysis. *J. Women Polit. Policy* **35**, 315–348 (2014).
73. Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C. & van Anders, S. M. The future of sex and gender in psychology: five challenges to the gender binary. *Am. Psychol.* **74**, 171–193 (2019).

Acknowledgements Early versions of this work were presented at the 2019 International Conference on Computational Social Science (IC2S2), the 2019 Conference on Politics and Computational Social Science (PaCCS) and the 2020 annual meeting of the American Political Science Association (APSA). We are grateful to the New York University Social Media and Political Participation (SMaPP) lab, the Stanford Internet Observatory and the Stanford Social Media Lab for feedback, and to Muhammad Ahmad Bashir for development on the 2018 extension. This research was supported in part by the Democracy Fund, the William and Flora Hewlett Foundation and the National Science Foundation (IIS-1910064).

Author contributions R.E.R., C.W. and D.L. conceived of the research. K.O., C.W., D.L. and R.E.R. contributed to survey design. R.E.R. built the 2020 data collection instrument. J.G. designed the multivariate regression analysis. R.E.R. and J.G. analysed the data and R.E.R. wrote the paper with D.J.R., J.G., K.O., C.W. and D.L. All authors approved the final manuscript.

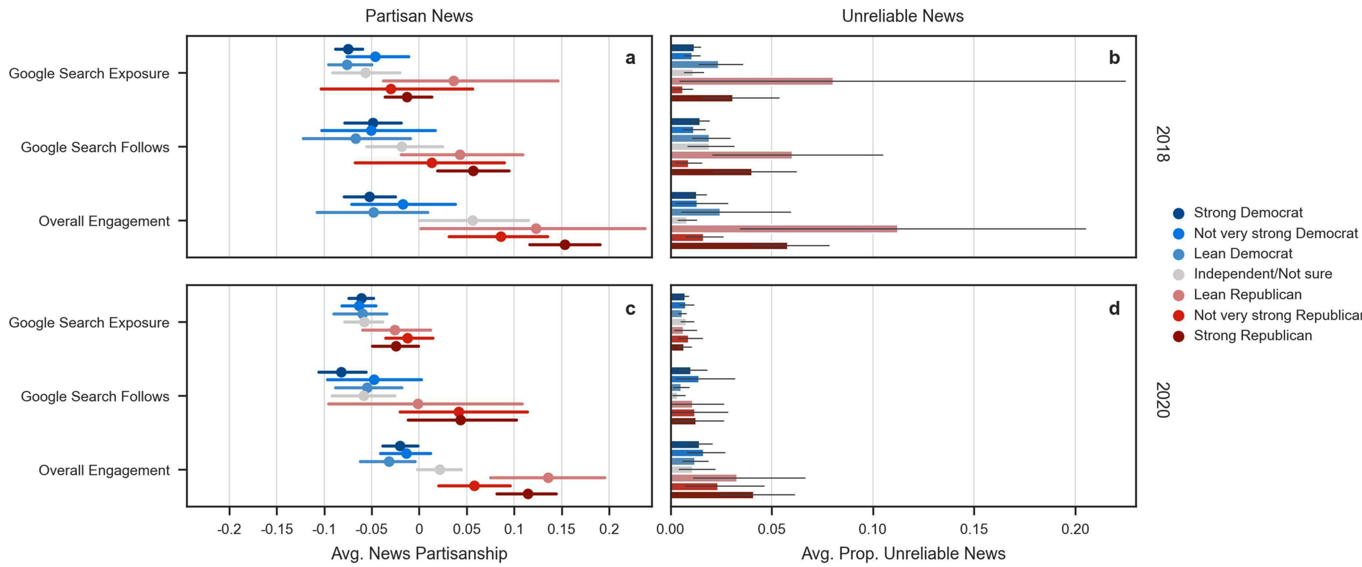
Competing interests The authors declare no competing interests.

Additional information

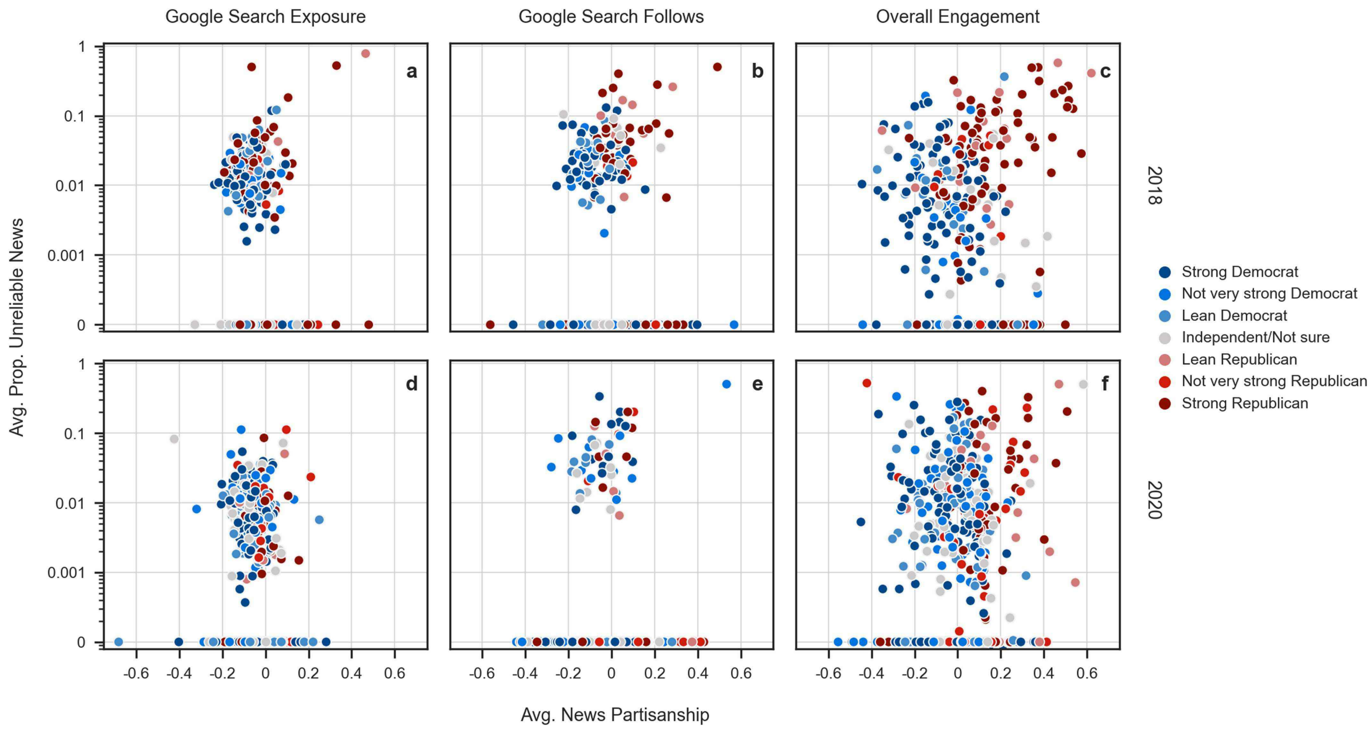
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-06078-5.

Correspondence and requests for materials should be addressed to Ronald E. Robertson.

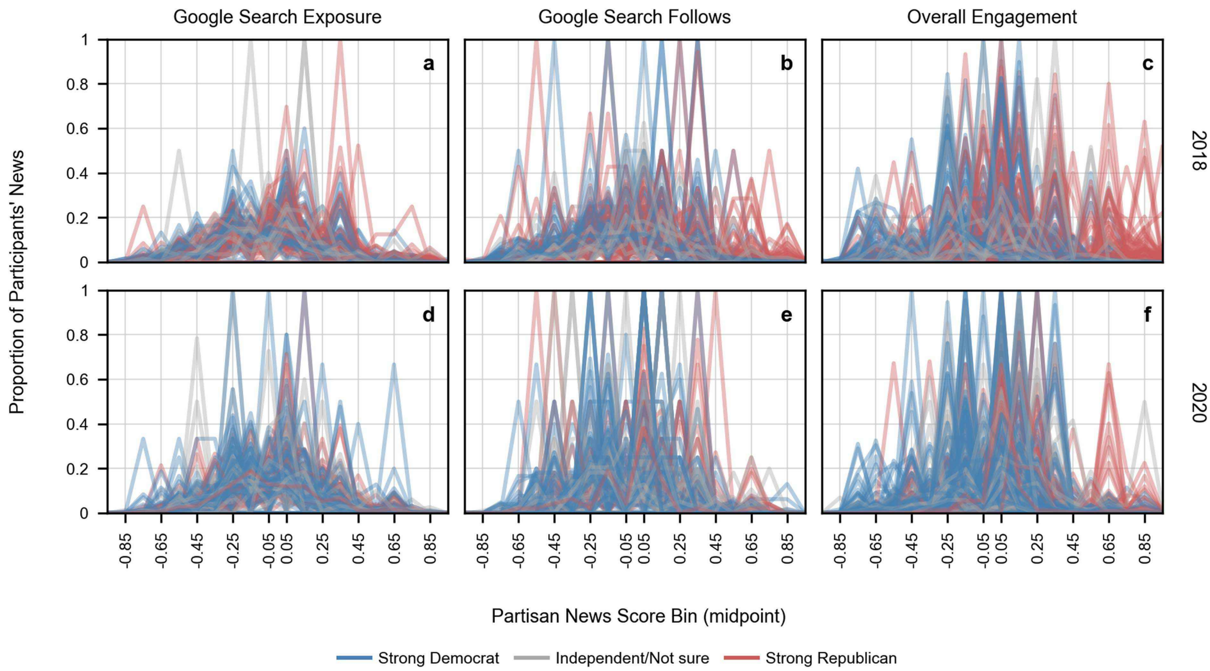
Peer review information *Nature* thanks Homa Hosseinmardi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | Strong partisans are exposed to similar rates of partisan and unreliable news, but asymmetrically follow and engage with such news. This figure complements Fig. 1 in the main text by displaying, for all 7-point PID groups, average exposure, follows and overall engagement with partisan (**a, c**) and unreliable news (**b, d**) by study wave and 7-point PID clustered at the participant-level. Data are presented as participant-level means grouped by 7-point PID in each subplot, all error bars indicate 95% confidence intervals (CI), and results from bivariate tests of differences in partisan and unreliable news by 7-point PID are available in Extended Data Table 2. A score of zero does not imply neutrality in the scores we used, so left-of-zero scores do not imply a left-leaning bias (Methods, 'Partisan News Scores').



Extended Data Fig. 2 | Partisans who engage with more identity-congruent news also tend to engage with more unreliable news. This figure complements Fig. 3 in the main text by displaying all 7-point PID groups, highlighting the relationship between partisan and unreliable news for participants' exposure on Google Search (**a, d**), follows from Google Search (**b, e**), and overall engagement (**c, f**). These subplots show that the relationship between partisan and unreliable news varies across data types, and within data types when taking partisan identity into account (Extended Data Table 3).



Extended Data Fig. 3 | Partisan news distributions at the participant level for each dataset and study wave. Each line represents the distribution of partisan news sources that a single participant was exposed to in their Google Search results (**a, d**), followed from those results (**b, e**), or engaged with overall (**c, f**). Partisan news scores have been binned in 0.1 point intervals (e.g. −1 to −0.9, −0.9 to −0.8, etc.) along the x-axis, with tick labels showing the midpoints of those bins.

Extended Data Table 1 | Descriptive counts for each exposure, follows, and overall engagement dataset

Year	Type	Data	Days Observed	Participants	Searches	URLs	News	Unreliable
2018	Google Search Exposure	Google SERPs	74	275	102,114	1,245,155	215,699	3,932
	Google Search Follows	Google History	254	262	-	279,680	22,946	661
	Overall Engagement	Browser History	253	333	-	14,677,297	994,032	29,081
		Google History	254	271	-	4,807,758	405,596	16,754
2020	Google Search Exposure	Google SERPs	203	459	226,035	3,654,829	586,803	5,184
	Google Search Follows	Tab Activity	203	418	-	69,023	8,125	112
	Overall Engagement	Browser History	291	688	-	31,202,830	1,862,011	13,209
		Tab Activity	205	597	-	20,260,394	1,897,933	11,880

Descriptive counts of the number of days we observed participants for, the number of participants, the number of searches they conducted, and the number of URLs, news, and unreliable news they encountered in each of our datasets. Days observed is the number of days between the first and last timestamp across all participants in each dataset. For data collected from existing APIs or interfaces, such as Browser History and Google History, we observed a greater number of days because they allow collection of data prior to the installation of the extension. In contrast, because we built the data collection instruments for Google SERPs and Tab Activity, these can only document participant behavior after the extension has been installed.

Extended Data Table 2 | Kruskal-Wallis *H* tests by study wave, data type, data source, and user grouping

				Partisanship			Unreliable		
Year	Type	Data	Group	N	H	P	N	H	P
2018	Google Search Exposure	Google SERPs	7-Point Party ID	270	23.25	0.001***	275	11.15	0.084
			Age Group	270	12.64	0.005**	275	0.71	0.871
	Google Search Follows	Google History	7-Point Party ID	247	36.27	0.000***	262	5.21	0.517
			Age Group	247	6.80	0.079	262	1.85	0.605
	Overall Engagement	Browser History	7-Point Party ID	319	72.22	0.000***	333	34.06	0.000***
			Age Group	319	23.23	0.000***	333	19.01	0.000***
		Google History	7-Point Party ID	270	66.86	0.000***	271	16.26	0.012*
			Age Group	270	11.78	0.008**	271	2.45	0.484
2020	Google Search Exposure	Google SERPs	7-Point Party ID	453	18.97	0.004**	459	4.05	0.670
			Age Group	453	29.48	0.000***	459	1.22	0.748
	Google Search Follows	Tab Activity	7-Point Party ID	329	28.01	0.000***	418	2.85	0.827
			Age Group	329	24.25	0.000***	418	0.57	0.903
	Overall Engagement	Browser History	7-Point Party ID	655	86.53	0.000***	688	18.17	0.006**
			Age Group	655	29.28	0.000***	688	32.01	0.000***
		Tab Activity	7-Point Party ID	546	57.72	0.000***	597	19.18	0.004**
			Age Group	546	27.54	0.000***	597	34.76	0.000***

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

For our key variables—partisan and unreliable news—we initially checked for statistically significant differences by partisan identity and age group using the Kruskal-Wallis *H* test, a non-parametric test of differences among three or more groups. We used this nonparametric test to account for the heterogeneity we observed in the distribution of participant-level averages (Methods, 'Descriptive Analysis). The Partisanship column contains comparisons of exposure, follows, and overall engagement with partisan news, and the Unreliable column compares differences in exposure, follows, and overall engagement with unreliable news. Differences in average partisan and unreliable news for all 7-point partisan identity groups are available in Extended Data Fig. 1. These results show a general absence of statistically significant differences in participants' exposure and follows to unreliable news via Google Search in both study waves.

Extended Data Table 3 | Spearman's ρ rank correlations comparing average news partisanship and proportion of unreliable news

			Strong Democrat			Independent/ Not sure			Strong Republican		
Year	Type	Data	N	ρ	P	N	ρ	P	N	ρ	P
2018	Google Search Exposure	Google SERPs	98	-0.133	0.956	28	0.058	1.000	74	-0.030	1.000
	Google Search Follows	Google History	90	-0.273	0.064	25	0.218	1.000	63	0.029	1.000
	Overall Engagement	Browser History	115	-0.380	0.000***	30	-0.093	1.000	89	0.336	0.008**
		Google History	98	-0.192	0.346	27	-0.059	0.890	70	0.377	0.009**
2020	Google Search Exposure	Google SERPs	130	-0.149	0.627	74	-0.140	1.000	50	-0.109	1.000
	Google Search Follows	Tab Activity	97	0.205	0.311	54	-0.080	1.000	36	-0.103	1.000
	Overall Engagement	Browser History	191	-0.237	0.007**	104	0.027	1.000	73	0.223	0.345
		Tab Activity	154	-0.140	0.495	88	-0.013	1.000	63	0.407	0.006**

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Spearman rank correlations comparing participants' average news partisanship and proportion of unreliable news by year, data type, data source, and partisan identity. Similar to our use of the Kruskal-Wallis H test, we used Spearman's rank correlation coefficient (ρ) to evaluate the relationship between partisan and unreliable news because it is a nonparametric test that helps to account for the heterogeneity we observed in the distribution of participant-level averages (Supplementary Information Table S5). Using a two-sided P value, we found no significant correlations between unreliable and partisan news in any dataset when considering all participants, except for Google Search exposure in 2020 ($N = 453$, $\rho = -0.162$, $P < 0.001$). However, when we examined these correlations within each 7-point partisan ID group (adjusting P values for multiple hypothesis testing using the Holm-Bonferroni method), the correlation for exposure was not significant for any group in either study wave, and significant negative correlations emerged for strong Democrats while significant positive correlations emerged for strong Republicans, suggesting a Simpson's paradox. We show the relationship between partisan and unreliable news for all 7-point partisan identities in Extended Data Fig. 2.

Extended Data Table 4 | Main multivariate regression results for partisan news in 2018

Type	Group	Label	N	Demographics β (95% CI)	t	P	Demographics + Query β (95% CI)	t	P
Google Search Exposure	7-Point Party ID	Strong Dem.	98	-0.008 (-0.051, 0.035)	-0.35	0.730	0.001 (-0.039, 0.040)	0.03	0.976
		Not very strong Dem.	26	0.013 (-0.039, 0.065)	0.50	0.878	0.027 (-0.021, 0.075)	1.12	0.527
		Lean Dem.	23	-0.014 (-0.069, 0.042)	-0.49	1.000	0.001 (-0.050, 0.051)	0.02	1.000
		Lean Rep.	11	0.097 (0.028, 0.165)	2.78	0.012*	0.093 (0.031, 0.156)	2.94	0.007**
		Not very strong Rep.	10	0.017 (-0.054, 0.088)	0.47	0.640	0.012 (-0.052, 0.077)	0.37	0.712
		Strong Rep.	74	0.027 (-0.016, 0.071)	1.23	0.439	0.031 (-0.009, 0.071)	1.53	0.255
	Age Group	25-44	93	0.067 (0.019, 0.116)	2.74	0.013*	0.085 (0.038, 0.131)	3.58	0.001***
		45-64	108	0.069 (0.020, 0.117)	2.80	0.011*	0.081 (0.034, 0.128)	3.41	0.002**
		65+	49	0.094 (0.040, 0.147)	3.45	0.001**	0.115 (0.064, 0.166)	4.46	0.000***
Google Search Follows	7-Point Party ID	Strong Dem.	90	-0.010 (-0.076, 0.055)	-0.31	0.806	-0.006 (-0.072, 0.060)	-0.17	0.863
		Not very strong Dem.	26	-0.018 (-0.096, 0.061)	-0.45	0.655	0.013 (-0.063, 0.090)	0.34	0.736
		Lean Dem.	19	-0.029 (-0.117, 0.059)	-0.65	1.000	-0.024 (-0.111, 0.062)	-0.55	0.979
		Lean Rep.	12	0.068 (-0.032, 0.169)	1.34	0.181	0.100 (0.003, 0.196)	2.03	0.088
		Not very strong Rep.	12	0.029 (-0.069, 0.127)	0.58	1.000	0.032 (-0.071, 0.136)	0.62	1.000
		Strong Rep.	63	0.073 (0.006, 0.141)	2.14	0.067	0.085 (0.018, 0.151)	2.50	0.026*
	Age Group	25-44	90	0.056 (-0.018, 0.131)	1.50	0.272	0.069 (-0.006, 0.144)	1.81	0.143
		45-64	95	0.071 (-0.003, 0.145)	1.88	0.123	0.095 (0.020, 0.171)	2.48	0.028*
		65+	44	0.069 (-0.013, 0.152)	1.66	0.198	0.090 (0.007, 0.174)	2.14	0.068
Overall Engagement	7-Point Party ID	Strong Dem.	115	-0.097 (-0.164, -0.029)	-2.81	0.011*	-	-	-
		Not very strong Dem.	29	-0.065 (-0.149, 0.019)	-1.53	0.254	-	-	-
		Lean Dem.	24	-0.100 (-0.190, -0.009)	-2.16	0.063	-	-	-
		Lean Rep.	15	0.081 (-0.022, 0.183)	1.55	0.121	-	-	-
		Not very strong Rep.	17	0.035 (-0.063, 0.133)	0.70	0.967	-	-	-
		Strong Rep.	89	0.099 (0.029, 0.168)	2.80	0.005**	-	-	-
	Age Group	25-44	110	0.085 (0.010, 0.161)	2.23	0.053	-	-	-
		45-64	122	0.111 (0.035, 0.187)	2.88	0.009**	-	-	-
		65+	64	0.150 (0.069, 0.231)	3.63	0.001***	-	-	-

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Regression coefficients (β), 95% confidence intervals (95% CI), t statistics, and P values for the associations between partisan news and our main independent variables (partisan identity and age) we found for each data type in 2018. This table shows results from our two primary model specifications, one that controlled for additional demographics ("Demographics") and one that controlled for both demographics and the text of participants' search queries ("Demographics + Query"). For 7-point partisan IDs, coefficients are relative to independents, and for age groups, coefficients are relative to the 18–24 age group. All P values have been corrected for multiple comparisons across outcomes using the Holm-Bonferroni method. Additional details on the design of our partisan news regressions are available in Methods ('Multivariate Regressions'), and detailed results for each specification we ran on partisan news in 2018 are available in Supplementary Information for Google Search exposure (Table S7), Google Search follows (Table S8), and overall engagement (Table S9).

Extended Data Table 5 | Main multivariate regression results for partisan news in 2020

				Demographics			Demographics + Query		
Type	Group	Label	N	β (95% CI)	t	P	β (95% CI)	t	P
Google Search Exposure	7-Point Party ID	Strong Dem.	130	0.007 (-0.019, 0.032)	0.51	1.000	0.004 (-0.021, 0.028)	0.30	1.000
		Not very strong Dem.	72	0.003 (-0.024, 0.031)	0.22	1.000	0.004 (-0.023, 0.030)	0.26	1.000
		Lean Dem.	69	0.002 (-0.025, 0.030)	0.16	1.000	0.002 (-0.025, 0.028)	0.12	1.000
		Lean Rep.	19	0.017 (-0.027, 0.060)	0.76	0.901	0.023 (-0.018, 0.064)	1.09	0.550
		Not very strong Rep.	39	0.041 (0.008, 0.074)	2.47	0.028*	0.036 (0.004, 0.068)	2.24	0.051
		Strong Rep.	50	0.029 (-0.002, 0.060)	1.82	0.139	0.016 (-0.014, 0.046)	1.04	0.601
	Age Group	25-44	166	0.018 (-0.009, 0.045)	1.30	0.363	0.013 (-0.013, 0.040)	0.99	0.322
		45-64	151	0.026 (-0.002, 0.054)	1.80	0.145	0.018 (-0.009, 0.045)	1.32	0.378
		65+	86	0.048 (0.017, 0.078)	3.05	0.005**	0.036 (0.006, 0.066)	2.38	0.036*
Google Search Follows	7-Point Party ID	Strong Dem.	97	-0.017 (-0.067, 0.033)	-0.67	1.000	-0.015 (-0.066, 0.035)	-0.59	1.000
		Not very strong Dem.	50	0.008 (-0.047, 0.064)	0.30	1.000	0.011 (-0.046, 0.067)	0.37	1.000
		Lean Dem.	51	-0.004 (-0.059, 0.051)	-0.15	1.000	-0.004 (-0.059, 0.051)	-0.14	1.000
		Lean Rep.	14	0.023 (-0.062, 0.108)	0.53	1.000	0.039 (-0.048, 0.126)	0.88	0.762
		Not very strong Rep.	27	0.081 (0.015, 0.147)	2.41	0.033*	0.081 (0.014, 0.148)	2.38	0.036*
		Strong Rep.	36	0.072 (0.010, 0.135)	2.27	0.048*	0.077 (0.012, 0.143)	2.32	0.042*
	Age Group	25-44	116	0.027 (-0.027, 0.080)	0.97	0.666	0.027 (-0.028, 0.082)	0.98	0.496
		45-64	109	0.049 (-0.006, 0.103)	1.76	0.160	0.047 (-0.009, 0.103)	1.65	0.105
		65+	66	0.106 (0.047, 0.165)	3.52	0.001**	0.092 (0.030, 0.153)	2.93	0.007**
Overall Engagement	7-Point Party ID	Strong Dem.	191	-0.031 (-0.065, 0.003)	-1.80	0.145	-	-	-
		Not very strong Dem.	106	-0.032 (-0.069, 0.005)	-1.68	0.186	-	-	-
		Lean Dem.	87	-0.054 (-0.093, -0.015)	-2.73	0.013*	-	-	-
		Lean Rep.	35	0.094 (0.041, 0.147)	3.48	0.001**	-	-	-
		Not very strong Rep.	59	0.026 (-0.018, 0.070)	1.16	0.245	-	-	-
		Strong Rep.	73	0.080 (0.038, 0.122)	3.71	0.000***	-	-	-
	Age Group	25-44	232	0.021 (-0.017, 0.058)	1.07	0.285	-	-	-
		45-64	226	0.053 (0.014, 0.092)	2.68	0.007**	-	-	-
		65+	128	0.068 (0.026, 0.110)	3.18	0.003**	-	-	-

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Regression coefficients (β), 95% confidence intervals (95% CI), t statistics, and P values for the associations between partisan news and our main independent variables (partisan identity and age) that we found for each data type in 2020. This table shows results from our two primary model specifications, one that controlled for additional demographics ("Demographics") and one that controlled for both demographics and the text of participants' search queries ("Demographics + Query"). For 7-point partisan IDs, coefficients are relative to independents, and for age groups, coefficients are relative to the 18–24 age group. All P values have been corrected for multiple comparisons across outcomes using the Holm-Bonferroni method. Additional details on the design of our partisan news regressions are available in Methods ('Multivariate Regressions'), and detailed results for each specification we ran on partisan news in 2020 are available in Supplementary Information for Google Search exposure (Table S10), Google Search follows (Table S11), and overall engagement (Table S12).

Extended Data Table 6 | Main multivariate regression results for unreliable news in 2018

				Demographics			Demographics + Query		
Type	Group	Label	N	β (95% CI)	t	P	β (95% CI)	t	P
Google Search Exposure	7-Point Party ID	Strong Dem.	99	-0.391 (-0.877, 0.094)	-1.59	0.227	-0.337 (-0.775, 0.102)	-1.51	0.263
		Not very strong Dem.	27	-0.229 (-0.812, 0.353)	-0.78	0.878	-0.265 (-0.786, 0.257)	-1.00	0.527
		Lean Dem.	23	0.025 (-0.562, 0.613)	0.08	1.000	0.039 (-0.496, 0.574)	0.14	1.000
		Lean Rep.	11	0.216 (-0.556, 0.989)	0.55	0.582	-0.103 (-0.843, 0.638)	-0.27	0.785
		Not very strong Rep.	10	-0.732 (-1.572, 0.108)	-1.72	0.174	-0.504 (-1.262, 0.255)	-1.31	0.384
		Strong Rep.	77	0.192 (-0.293, 0.677)	0.78	0.439	0.124 (-0.314, 0.562)	0.56	0.578
	Age Group	25-44	94	0.128 (-0.354, 0.609)	0.52	0.602	0.169 (-0.285, 0.623)	0.73	0.463
		45-64	111	-0.024 (-0.507, 0.460)	-0.10	0.924	0.060 (-0.402, 0.522)	0.25	0.799
		65+	50	0.370 (-0.172, 0.912)	1.34	0.180	0.386 (-0.127, 0.898)	1.48	0.140
Google Search Follows	7-Point Party ID	Strong Dem.	95	-0.279 (-0.934, 0.377)	-0.84	0.806	-0.449 (-1.100, 0.202)	-1.36	0.350
		Not very strong Dem.	28	-0.427 (-1.206, 0.353)	-1.08	0.564	-0.640 (-1.397, 0.118)	-1.66	0.195
		Lean Dem.	20	-0.269 (-1.086, 0.549)	-0.65	1.000	-0.276 (-1.060, 0.509)	-0.69	0.979
		Lean Rep.	13	0.796 (-0.119, 1.711)	1.71	0.175	0.675 (-0.219, 1.569)	1.49	0.138
		Not very strong Rep.	13	-0.345 (-1.361, 0.671)	-0.67	1.000	-0.286 (-1.306, 0.735)	-0.55	1.000
		Strong Rep.	68	0.368 (-0.292, 1.029)	1.10	0.273	0.193 (-0.457, 0.842)	0.58	0.559
	Age Group	25-44	94	0.307 (-0.342, 0.957)	0.93	0.352	0.252 (-0.387, 0.890)	0.78	0.438
		45-64	100	0.258 (-0.402, 0.919)	0.77	0.441	0.215 (-0.436, 0.865)	0.65	0.516
		65+	49	0.423 (-0.319, 1.165)	1.12	0.262	0.304 (-0.411, 1.019)	0.84	0.403
Overall Engagement	7-Point Party ID	Strong Dem.	120	-0.060 (-0.834, 0.713)	-0.15	0.878	-	-	-
		Not very strong Dem.	32	-0.067 (-0.992, 0.858)	-0.14	0.887	-	-	-
		Lean Dem.	24	0.315 (-0.672, 1.303)	0.63	0.531	-	-	-
		Lean Rep.	15	1.894 (0.817, 2.970)	3.46	0.001**	-	-	-
		Not very strong Rep.	18	0.372 (-0.714, 1.458)	0.67	0.967	-	-	-
		Strong Rep.	91	1.268 (0.497, 2.039)	3.23	0.003**	-	-	-
	Age Group	25-44	117	-0.226 (-1.042, 0.589)	-0.55	0.585	-	-	-
		45-64	126	0.354 (-0.455, 1.163)	0.86	0.389	-	-	-
		65+	66	0.470 (-0.389, 1.330)	1.08	0.283	-	-	-

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Regression coefficients (β), 95% confidence intervals (95% CI), t statistics, and P values for the associations between unreliable news and our main independent variables (partisan identity and age) that we found for each data type in 2018. This table shows results from our two primary model specifications, one that controlled for additional demographics ("Demographics") and one that controlled for both demographics and the text of participants' search queries ("Demographics + Query"). For 7-point partisan IDs, coefficients are relative to independents, and for age groups, coefficients are relative to the 18–24 age group. All P values have been corrected for multiple comparisons across outcomes using the Holm-Bonferroni method. Additional details on the design of our unreliable news regressions are available in Methods ('Multivariate Regressions'), and detailed results for each specification we ran on unreliable news in 2018 are available in Supplementary Information for Google Search exposure (Table S13), Google Search follows (Table S14), and overall engagement (Table S15).

Extended Data Table 7 | Main multivariate regression results for unreliable news in 2018

				Demographics			Demographics + Query		
Type	Group	Label	N	β (95% CI)	t	P	β (95% CI)	t	P
Google Search Exposure	7-Point Party ID	Strong Dem.	130	0.003 (-0.374, 0.379)	0.01	1.000	0.018 (-0.358, 0.394)	0.09	1.000
		Not very strong Dem.	75	0.100 (-0.312, 0.512)	0.48	1.000	0.092 (-0.320, 0.503)	0.44	1.000
		Lean Dem.	70	0.010 (-0.399, 0.420)	0.05	1.000	0.003 (-0.409, 0.414)	0.01	1.000
		Lean Rep.	19	-0.188 (-0.872, 0.496)	-0.54	0.901	-0.209 (-0.896, 0.478)	-0.60	0.550
		Not very strong Rep.	39	0.052 (-0.454, 0.558)	0.20	0.839	0.152 (-0.356, 0.660)	0.59	0.558
		Strong Rep.	51	-0.098 (-0.576, 0.380)	-0.40	0.687	0.071 (-0.414, 0.557)	0.29	0.773
	Age Group	25-44	169	0.270 (-0.127, 0.667)	1.34	0.363	0.292 (-0.109, 0.693)	1.43	0.306
		45-64	152	0.157 (-0.255, 0.568)	0.75	0.455	0.230 (-0.188, 0.649)	1.08	0.378
		65+	88	0.104 (-0.345, 0.552)	0.45	0.650	0.139 (-0.325, 0.602)	0.59	0.557
Google Search Follows	7-Point Party ID	Strong Dem.	120	-0.060 (-1.084, 0.963)	-0.12	1.000	0.224 (-0.861, 1.310)	0.41	1.000
		Not very strong Dem.	63	-0.035 (-1.113, 1.043)	-0.06	1.000	0.105 (-1.013, 1.224)	0.19	1.000
		Lean Dem.	63	0.145 (-0.910, 1.200)	0.27	1.000	0.181 (-0.939, 1.301)	0.32	1.000
		Lean Rep.	18	0.049 (-1.546, 1.644)	0.06	1.000	0.166 (-1.503, 1.835)	0.20	0.845
		Not very strong Rep.	37	0.678 (-0.543, 1.898)	1.09	0.276	1.024 (-0.282, 2.330)	1.54	0.124
		Strong Rep.	45	0.285 (-0.954, 1.523)	0.45	0.652	0.429 (-0.932, 1.789)	0.62	0.536
	Age Group	25-44	154	0.464 (-0.693, 1.621)	0.79	0.666	0.730 (-0.511, 1.971)	1.16	0.496
		45-64	138	0.859 (-0.292, 2.010)	1.47	0.160	1.258 (-0.014, 2.529)	1.94	0.105
		65+	83	0.942 (-0.274, 2.158)	1.52	0.129	1.201 (-0.150, 2.553)	1.75	0.081
Overall Engagement	7-Point Party ID	Strong Dem.	199	0.232 (-0.283, 0.747)	0.88	0.378	-	-	-
		Not very strong Dem.	112	0.255 (-0.316, 0.826)	0.88	0.381	-	-	-
		Lean Dem.	92	0.160 (-0.426, 0.747)	0.54	0.592	-	-	-
		Lean Rep.	36	0.644 (-0.137, 1.424)	1.62	0.106	-	-	-
		Not very strong Rep.	60	0.608 (-0.051, 1.267)	1.81	0.141	-	-	-
		Strong Rep.	76	1.141 (0.529, 1.752)	3.66	0.000***	-	-	-
	Age Group	25-44	247	0.523 (-0.088, 1.134)	1.68	0.186	-	-	-
		45-64	237	1.014 (0.399, 1.630)	3.23	0.003**	-	-	-
		65+	128	1.067 (0.413, 1.721)	3.20	0.003**	-	-	-

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Regression coefficients (β), 95% confidence intervals (95% CI), t statistics, and P values for the associations between unreliable news and our main independent variables (partisan identity and age) that we found for each data type in 2020. This table shows results from our two primary model specifications, one that controlled for additional demographics ("Demographics") and one that controlled for both demographics and the text of participants' search queries ("Demographics + Query"). For 7-point partisan IDs, coefficients are relative to independents, and for age groups, coefficients are relative to the 18–24 age group. All P values have been corrected for multiple comparisons across outcomes using the Holm-Bonferroni method. (β)Additional details on the design of our unreliable news regressions are available in Methods ('Multivariate Regressions'), and detailed results for each specification we ran on unreliable news in 2020 are available in Supplementary Information for Google Search exposure (Table S16), Google Search follows (Table S17), and overall engagement (Table S18).

nature portfolio

Corresponding author(s): Ronald E. Robertson

Last updated by author(s): Dec 21, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The data for this study were collected using custom browser extensions written in JavaScript and using the WebExtension framework for cross-browser compatibility. The source code of the extensions used in 2018 and 2020 will be posted upon publication.
Data analysis	Analyses were performed with Python nalyses were performed with Python v3.10.4, pandas v1.4.3, scipy v1.8.1, Spark v3.1, and R v4.1.The parser we used to extract the URLs our participants were exposed to is available at https://github.com/gitronald/WebSearcher.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Due to privacy concerns and IRB limitations, aggregated data are available upon request, but individual data cannot be shared. The domain scores and classifications we used are available at https://github.com/gitronald/domains, but the NewsGuard classifications are not included due to their proprietary nature.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

This study involves a quantitative observational examination of digital trace data collected from participants in the U.S. around the time of the 2018 and 2020 U.S. elections.

Research sample

Participants were recruited through YouGov in 2018, and via PureSpectrum in 2020 as part of the "COVID States Project" (see https://covidstates.org). For our recruitment via YouGov in 2018 we requested an oversampling of strong partisans to ensure statistical leverage for our main comparisons. For 2020, our recruitment via PureSpectrum used quota sampling based on state-by-state benchmarks for race, gender, age, and location, but a non-random subset of these respondents opted-in to installing the browser extension. Compared to national estimates, our 2018 sample had fewer people aged 18-24, a greater proportion of strong partisans (indicating that our oversampling worked), and more participants who identified as White, male, and college-educated (see Table S1). Among our 2020 sample, the differences in age and partisan identity groups were closer to national estimates than in 2018, and a greater proportion of participants identified as female (see Table S2).

Sampling strategy

We used the sample pool provided by YouGov in 2018, and recruited subjects via PureSpectrum in 2020 using the quota sampling described above. Sample sizes were limited due to both recruitment costs and optional extension installations. We provide sample details in Supplementary Information Tables S1 and S2.

We did not conduct a power analysis prior to either study wave due to a combination of budget constraints, uncertainty around how many survey participants would opt-in to installing our browser extension (Supplementary Information S1.1), and uncertainty around potential effect sizes. However, the consistency of our main results across both study waves suggests that our sample sizes were sufficiently powered.

Data collection

Digital trace data were collected from participants via custom browser extensions that we built for Chrome and Firefox. Aside from the established domain-level metrics we describe in Methods, we collected all data used in this study directly from our participants in real time, or on a fixed schedule, depending on the dataset. The Google History and Browser History data we collected may be considered preexisting—because both are collected by third-party services that operate independently of our study—from the perspective of an individual participant or the service. For Google History, that data collection is managed by Google and its account holders, and for Browser History, it is managed by Chrome or Firefox and their web browser users. We provide additional details on each dataset we collected in Methods (under Exposure and Engagement Datasets).

Before installing the browser extension, participants were informed of all data types that the browser extension collected, and of our broad aim to understand how people encounter online news. The protocol and informed consent language we used while recruiting subjects was transparent about what the extension collected, and both studies were approved by the IRB at Northeastern University (#18-10-03 for 2018, and #20-03-04 for 2020). Participants had anonymous identifiers, but our analysis was not otherwise blinded.

Timing

The first survey was fielded between October 18 and October 24 of 2018 and the second was fielded between June 12, 2020 and January 6, 2021.

Data exclusions

We identified and removed consecutive visits to the same web page that occurred within one second, keeping only the first instance. Consecutive visits to the same web page are often present in such data, and can occur for a variety of reasons, such as refreshing a stalled page, or website-specific idiosyncrasies in page loading. We also filtered out participants whose participation window---the duration between their first and last observed behavior---was less than 10 days, which suggests that they installed and then quickly uninstalled the extension. More details and exact numbers are provided in the main text and the Methods section.

Non-participation

In 2018, 11.5% of the 3,106 (357) survey participants recruited via YouGov installed the extension and provided at least some data. In 2020, we offered participants recruited via PureSpectrum the opportunity to opt-in to installing the extension as part of a much larger survey-focused project (see https://covidstates.org), and 0.4% of the 189,711 (759) survey participants installed the extension and provided at least some data.

Randomization

Participants were not allocated into experimental groups, but were grouped by survey responses on key demographic variables (age and partisan identity).

When adjusting our regression results for multiple comparisons, we did not correct across model specifications, because different specifications within the same model are not independent, nor across data types (exposure/follows/engagement), because they are also not fully independent of one another. Since we conducted the 2020 study wave as a replication, which is another way of guarding against false discovery, we also did not correct across study waves. For outcomes (partisan and unreliable news), we have theoretical reasons to expect age to matter for unreliable news, but not for partisan news. However, to test the robustness of our results, we adjusted P values across outcomes for our primary models (Extended Data Tables 4-7) using the Holm-Bonferroni method.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Please see Supplementary Information Tables S1 and S2 for extended demographic details and comparisons. As mentioned above, our 2018 sample had fewer people aged 18-24, a greater proportion of strong partisans (indicating our oversampling worked), and more participants who identified as White, male, and college-educated compared to national estimates (see Table S1). For our 2020 sample, the differences in age and partisan identity groups were closer to national estimates than our 2018 sample was, and a greater proportion of participants identified as female (see Table S2).
Recruitment	As mentioned above, participants were recruited with the assistance of two different survey vendors: YouGov in 2018 and PureSpectrum in 2020. Self-selection for installing the browser extension is likely, suggested by the low opt-in rates. However, the consistency of our results in 2018 and 2020 suggests that our findings hold across substantively different samples.
Ethics oversight	As noted in the paper, both studies were approved by the IRB at Northeastern University (#18-10-03 for 2018, and #20-03-04 for 2020).

Note that full information on the approval of the study protocol must also be provided in the manuscript.