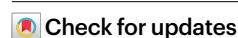


Sharing without clicking on news in social media

Received: 18 October 2023

Accepted: 17 October 2024

Published online: 19 November 2024



S. Shyam Sundar¹✉, Eugene Cho Snyder², Mengqi Liao³, Junjun Yin⁴,
Jinping Wang⁵ & Guangqing Chi⁴

Social media have enabled laypersons to disseminate, at scale, links to news and public affairs information. Many individuals share such links without first reading the linked information. Here we analysed over 35 million public Facebook posts with uniform resource locators shared between 2017 and 2020, and discovered that such ‘shares without clicks’ (SwoCs) constitute around 75% of forwarded links. Extreme and user-aligned political content received more SwoCs, with partisans engaging in it more than politically neutral users. In addition, analyses with 2,969 false uniform resource locators revealed higher shares and, hence, SwoCs by conservatives (76.94%) than liberals (14.25%), probably because, in our dataset, the vast majority (76–82%) of them originated from conservative news domains. Findings suggest that the virality of political content on social media (including misinformation) is driven by superficial processing of headlines and blurbs rather than systematic processing of core content, which has design implications for promoting deliberate discourse in the online public sphere.

Sharing is the hallmark of social media, leading to the virality of content and conferring status via metrics such as number of shares, reposts and retweets. In the realm of politics, this activity provides a voice to the masses by enabling them to share their political thought at scale. However, the absence of professional curation means that such sharing could also lead to the spread of misinformation and extreme political content. The rushed nature of online interactions has led many users to become less deliberate and more spontaneous in sharing social media content, resulting in the widespread phenomenon of ‘shares without clicks’ (SwoCs), wherein users forward uniform resource locators (URLs) of news stories and other public affairs information to others in their network without clicking through and reading it first. Different from shares with clicking, SwoCs are likely to be based on superficial cues (such as the headline or number of likes), without effortfully perusing the central contents of the news story. This suggests heuristic (as opposed to systematic) processing of news¹. Gabielkov et al.’s² analysis of Twitter activities during a month in 2015 suggested that 59% of the shared URLs for five major news domains were never clicked on,

providing aggregate evidence for this phenomenon. Holmström et al.³ reported similar findings by showing that a significant amount of URL links had more retweets than clicks, and revealed that 80% of all clicks occur within a day of a 5-day period, with links having more retweets than clicks increasing over time.

Research has revealed that users who snack on social media news with simple news blurbs (title, picture or news lead) acquire significantly less factual knowledge from their consumption of news across various digital platforms. Instead, it makes users feel more knowledgeable than they actually are (that is, an illusion of knowledge gain) about scientific information⁴ and political affairs⁵. This is because news headlines and blurbs are qualitatively different from full-length news articles in terms of framing, emphasis and tone⁶, true even for news coming from quality outlets such as *The New York Times*⁷. Furthermore, the increasing use of clickbait headlines to lure attention on social media can often be misleading^{8–10}. News headlines are shortcuts that provide readers who skim through news feeds with a different set of heuristic cues than those paying closer attention⁶.

¹Donald P. Bellisario College of Communications, Pennsylvania State University, University Park, PA, USA. ²Department of the Humanities and Social Sciences, New Jersey Institute of Technology, Newark, NJ, USA. ³Grady College of Journalism and Mass Communication, University of Georgia, Athens, GA, USA. ⁴Social Science Research Institute and Population Research Institute, Pennsylvania State University, University Park, PA, USA. ⁵College of Journalism and Communications, University of Florida, Gainesville, FL, USA. ✉e-mail: sss12@psu.edu

Table 1 | Data size and classification for each year

	2017 May–November	2018 midterm	2019 May–November	2020 presidential
Political classifier performance	Recall 0.8850	Recall 0.9225	Recall 0.8775	Recall 0.8775
	Precision 0.8850	Precision 0.9225	Precision 0.8775	Precision 0.8775
	F1 score 0.8847	F1 score 0.9156	F1 score 0.8833	F1 score 0.8816
Data size	4,195,082 URLs	7,068,001 URLs	10,503,168 URLs	13,494,326 URLs
Political content = 1	541,897 URLs (12.92%)	1,030,170 URLs (14.58%)	1,539,807 URLs (14.66%)	2,033,887 URLs (15.07%)
Non-political content = 0	3,200,575 URLs (76.29%)	5,848,741 URLs (82.75%)	7,121,032 URLs (67.80%)	9,301,016 URLs (68.93%)
Other category = 2	305,722 URLs (7.29%)	4816 URLs (0.01%)	1,592,641 URLs (15.16%)	1,801,750 URLs (13.35%)
Filtered data size	1,010,637 URLs (24.09% of all)	1,178,093 URLs (16.67% of all)	1,658,672 URLs (15.79% of all)	2,017,701 URLs (14.95% of all)
Political content = 1	176,286 URLs (17.44%)	205,565 URLs (17.45%)	269,127 URLs (16.23%)	372,279 URLs (18.45%)
Non-political content = 0	770,953 URLs (76.28%)	946,301 URLs (80.32%)	1,127,681 URLs (67.99%)	1,360,763 URLs (67.44%)
Other category = 2	46,428 URLs (4.59%)	543 URLs (0.05%)	219,656 URLs (13.24%)	226,726 URLs (11.24%)

‘Filtered dataset’ is a subset that includes only those URLs that met the minimum threshold (of 2,395 shares) required for overcoming noise due to ‘differential privacy’. ‘Other category’ refers to URLs with unclassifiable titles or blurbs based on our English-based categorization (for example, non-English blurbs) and those that returned error messages (for example, ‘Site Not Configured | 404 Not Found’), which we coded as 2. The URLs without any text (NaN) in both titles and main blurbs (for example, URLs from giphy.com, an online database that individuals use to share animated GIF files) were not assigned any political classifier codes. Thus, the sum of classified URLs is not equal to the entire dataset size. Please note that we used bold formatting for total (unfiltered and filtered) data sizes to enhance visibility of the representative sum of different content categories.

One reason for engaging in the seemingly indeliberate behaviour of sharing without clicking could be the information overload in personal and social media feeds, putting pressure on online users to be expedient and thereby leading them to rely on simple, often superficial, cues¹¹. For instance, extensive evidence shows how the sheer presence of bandwagon and social endorsement cues, such as number of likes¹² and comments¹², as well as the number of tweets and retweets¹³, can alter users’ credibility of social media content¹⁴ as well as their sharing/viewing intentions and behaviours^{15,16}. SwoC, however, is more consequential than likes and comments on a news article because it goes beyond signalling approval; users might be unwittingly contributing to bandwagon cascades and the spread of misinformation by sharing seemingly credible and highly endorsed content from third parties, often without personally reading, let alone verifying, that content.

To understand the specific cues that facilitate this quick but consequential action, we studied the sharing of political news on Facebook over a span of 4 years. We tested if SwoC is influenced by (1) the polarization level of the political content, (2) user alignment with the political content and (3) users’ political affinity. Furthermore, we explored if the SwoC phenomenon is specifically associated with false social media posts. First, polarized content can be easier for individuals to spot and distribute without deliberation. Hyper-partisan news is usually identified as being sensational and eye-catching given its reliance on intensified, exaggerated language and focus on dramatic events¹⁷. Clickbait headlines often communicate the essence of the news stories, thus eschewing the need for users to click on them, resulting in lower clicks¹⁸ but higher shares. Second, the ideological alignment of the political content can effectively serve as another heuristic cue. Users’ bias or preference towards ideologically consistent information on social media has been documented extensively^{19–21}. News consumers can easily detect partisanship from simply reading the headline²², and show a tendency to choose a source aligned with their own political stance²³, echoing a long line of research in political science demonstrating confirmation bias among partisans. Moreover, consuming ideologically similar information usually requires less cognitive effort²⁴, and users with lower analytical thinking abilities are known to have higher tendency to share hyper-partisan news headlines²³. This indicates how heuristic processing might contribute to SwoC. However, this tendency may not be equally true for all partisans. Research on Facebook use during the 2020 US presidential elections has noted asymmetry along ideological lines, with conservatives more likely to be exposed to

homogeneous information and misinformation²⁵. In the 2016 election season, Grinberg et al.²⁶ found that older, right-wing users tended to engage more intensively with false Twitter news. Selective exposure studies have found that right-wing participants are more likely to be influenced by social cues on the media interface when selecting news to read than left-wing users^{23,27}, which supports the literature of stronger conformity witnessed among conservatives^{28,29}. Thus, we examined if the intensity of this unique sharing behaviour (without clicking) varies on the basis of political leanings of the users. Lastly, we investigated whether the above SwoC patterns apply to false social media content, in light of the aforementioned concerns over misinformation spread, and also replicate at scale previous findings with small samples suggesting that users are more likely to spontaneously comment or share false (compared with true) social media content without reading it first³⁰.

Approach

Our findings are based on the analysis of SwoCs for all URLs shared on Facebook between 2017 and 2020 from the ‘Facebook URLs dataset’, offered by Social Science One in collaboration with Meta³¹. Specifically, we analysed 4,617 top-shared domains (for example, foxnews.com, cnn.com and nytimes.com) aggregated from the URL-level data across the 4-year period. To pinpoint patterns in political content, which is not detectable in a domain-level analysis, we also analysed URL data for a politically relevant timeframe in 2018, from May to November (that is, the 2018 midterm election season), as well as in 2020, from May to November (that is, the 2020 presidential election season). To offer comparison points, we also analysed data from May to November for 2017 and 2019. In all, we extracted users’ behavioural data pertaining to about 35 million URLs over 4 years (4,195,082 URLs from 2017, 7,068,001 URLs from 2018, 10,503,168 URLs from 2019 and 13,494,326 URLs from 2020). We created training data consisting of 8,000 URLs manually classified into political versus non-political, and then applied supervised machine learning to all URLs in our sample to identify political content. General data size and political content size for each year are summarized in Table 1.

Our main outcome variable, SwoC, is readily available in the dataset and refers to the “number of users who shared a post containing the URL but did not actually click on the link”³¹. For our three main predictors (that is, political affinity of content, political affinity of the user and the alignment level between those two), we calculated content affinity and alignment on the basis of the five-level user political affinity

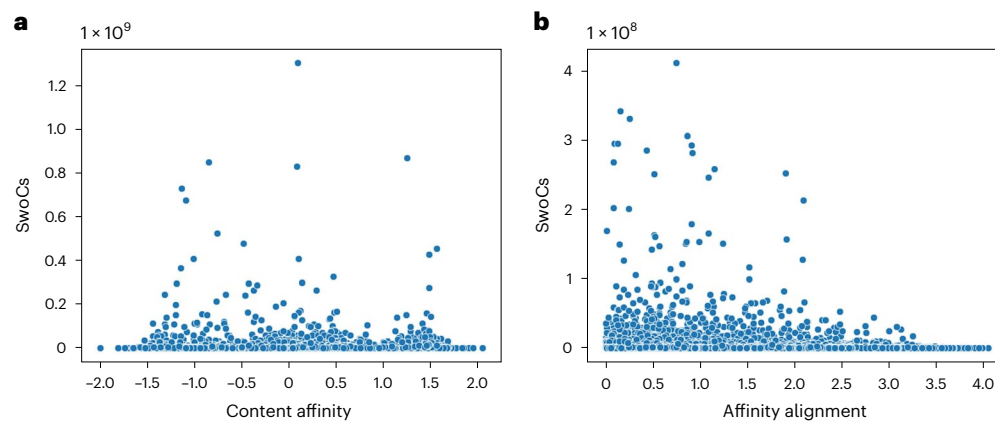


Fig. 1 | The relationship between political content affinity and affinity alignment with SwoCs of 4,617 top-shared domains. a, Content affinity ($b = -1,835,577.13$, s.e. 833,515.47, 95% confidence interval (CI) $-3,469,267.45$ to $-201,886.81$, $P = 0.028$), content affinity²: ($b = 2,412,064.41$, s.e. 847,008.07,

95% CI 751,928.59 to 4,072,200.23, $P = 0.004$), $R^2 = 0.002$. **b**, Alignment: $b = -1,053,149.64$, s.e. 76,048.43, 95% CI $-1,202,204.56$ to $-904,094.72$, $P < 0.001$, $R^2 = 0.008$.

scores (-2 = very liberal, -1 = moderately liberal, 0 = neutral, 1 = moderately conservative, 2 = very conservative) provided in the dataset. We triangulated our inferences by analysing data at two different units of analyses (domain level and URL level). First, we followed the recommendation of Buntain et al.³² to focus on popular domains with substantial amounts of sharing. This was to ensure that the added noise to the aggregated variables from ‘differential privacy’ (a privacy-preserving technique that offers a statistically acceptable range instead of actual amount of each user behaviour³¹) does not significantly affect the estimates in any statistical or practical sense. We added the URL-level analysis since domain-level data do not account for differences in content type among URLs (that is, political versus non-political) coming from a given domain (see ‘Source data characteristics’, ‘Construction of major variables’ and ‘Data preparation process’ sections in Methods for details of data characteristics and data preparation processes).

Results

Is polarized content associated with SwoCs?

To test the curvilinear effects of the political leaning of domains on SwoCs, we included the domain affinity score ($b = -1,835,577.13$, s.e. 833,515.47, $P = 0.03$) as well as its square ($b = 2,412,064.41$, s.e. 847,008.07, $P = 0.004$) in the regression model ($R^2 = 0.002$). Statistical indicators of the effects of squared domain affinity suggest a statistically significant pattern of more extreme domains (both far right and left) being shared without clicks (Fig. 1a). For URL-level analysis, we only selected political URLs with sufficient amount of shares, to overcome ‘differential privacy’ noise³². A clearer curvilinear visual data pattern emerged from URL-level data consistent across years, revealing that the more politically extreme the content (again, both liberal and conservative), the more it was shared without clicks, compared with neutral content (see Fig. 2a for 2017, Fig. 2c for 2018, Fig. 2e for 2019 and Fig. 2g

for 2020). When we ran identical analyses with non-political content, SwoCs were distributed evenly across content with different political affinity across 4 years (see Fig. 2b for 2017, Fig. 2d for 2018, Fig. 2f for 2019 and Fig. 2h for 2020). We acknowledge that non-political URLs are not likely to have political leaning, but we are reporting the analyses to offer a comparison point in relation to the patterns seen among the political URLs (see ‘URL-level analyses of the associations between major variables (content affinity, alignment and SwoCs) and fake news’ section in Methods for statistical details of the URL-level analysis).

Does politically aligned content attract more SwoCs?

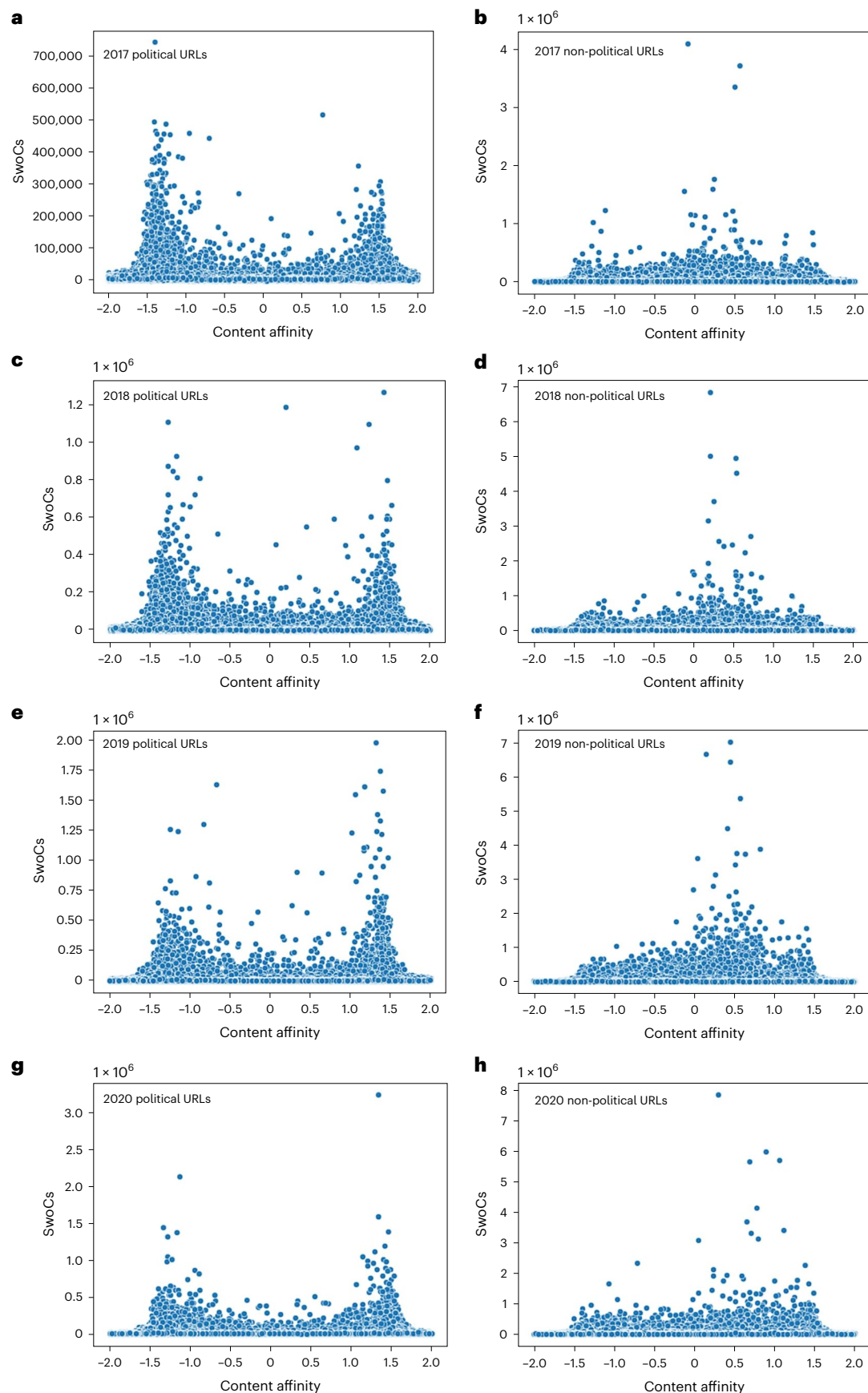
At the domain level, we found a clear pattern suggesting that the more politically aligned a domain is to the user (closer to 0), the more likely it is shared without being clicked upon first (Fig. 1b; $b = -1,053,149.64$, SE 76,048.43, $P < 0.001$) based on linear regression ($R^2 = 0.008$). As an independent measure, when we replaced our calculated content affinity scores with two alternative publicly available content affinity scores (that is, media bias ratings from AllSides.com and Robertson et al.³³), the affinity alignment effects were generally replicated at the domain level (see ‘Domain-level replication with external media bias ratings’ section in Methods for statistical details of the replication). Consistent with domain-level data, URLs more aligned with users’ political affinity were shared more without being clicked (see Fig. 3a for 2017, Fig. 3c for 2018, Fig. 3e for 2019 and Fig. 3g for 2020). When we ran the same analyses on non-political content, similar patterns emerged, but the effects were weaker (see Fig. 3b for 2017, Fig. 3d for 2018, Fig. 3f for 2019 and Fig. 3h for 2020).

Is political leaning of users associated with SwoCs?

When we compared SwoCs from 4,617 top domains by five political user affinity groups, the highest numbers (of SwoCs as well as overall

Fig. 2 | The relationship between content affinity and SwoCs of political and non-political URLs. a, 2017 May–November, political URLs: content affinity ($b = -967.63$, s.e. 31.25, 95% confidence interval (CI) $-1,028.88$ to -906.38 , $P < 0.001$), content affinity² ($b = 2,805.77$, s.e. 34.83, 95% CI 2,737.51 to 2,874.03, $P < 0.001$), $R^2 = 0.044$. **b**, 2017 May–November, non-political URLs: content affinity ($b = -29.38$, s.e. 21.15, 95% CI -70.84 to 12.08 , $P = 0.165$), content affinity² ($b = 335.55$, s.e. 20.86, 95% CI 294.67 to 376.43, $P < 0.001$), $R^2 = 0.0003$. **c**, 2018 midterm, political URLs: content affinity ($b = -754.20$, s.e. 42.57, 95% CI -837.62 to -670.77 , $P < 0.001$), content affinity² ($b = 3,630.11$, s.e. 44.70, 95% CI 3,542.50 to 3,717.73, $P < 0.001$), $R^2 = 0.034$. **d**, 2018 midterm, non-political URLs: content affinity ($b = 328.07$, s.e. 27.51, 95% CI 274.15 to 381.99, $P < 0.001$), content affinity² ($b = 282.71$, s.e. 27.19, 95% CI 229.41 to 336.01, $P < 0.001$), $R^2 = 0.0003$. **e**, 2019

May–November, political URLs: content affinity ($b = -71.96$, s.e. 46.70, 95% CI -163.48 to 19.57 , $P = 0.12$), content affinity² ($b = 3,745.15$, s.e. 48.46, 95% CI 3,650.18 to 3,840.12, $P < 0.001$), $R^2 = 0.023$. **f**, 2019 May–November, non-political URLs: content affinity ($b = 390.47$, s.e. 36.02, 95% CI 319.88 to 461.06, $P < 0.001$), content affinity² ($b = -81.58$, s.e. 35.94, 95% CI -152.03 to -11.14 , $P = 0.023$), $R^2 = 0.0001$. **g**, 2020 presidential, political URLs: content affinity ($b = 297.36$, s.e. 35.90, 95% CI 226.99 to 367.73, $P < 0.001$), content affinity² ($b = 4,196.67$, s.e. 37.41, 95% CI 4,123.36 to 4,269.98, $P < 0.001$), $R^2 = 0.035$. **h**, 2020 presidential, non-political URLs: content affinity ($b = 820.77$, s.e. 23.27, 95% CI 775.16 to 866.37, $P < 0.001$), content affinity² ($b = 1,388.28$, s.e. 23.01, 95% CI 1,343.18 to 1,433.38, $P < 0.001$), $R^2 = 0.004$.



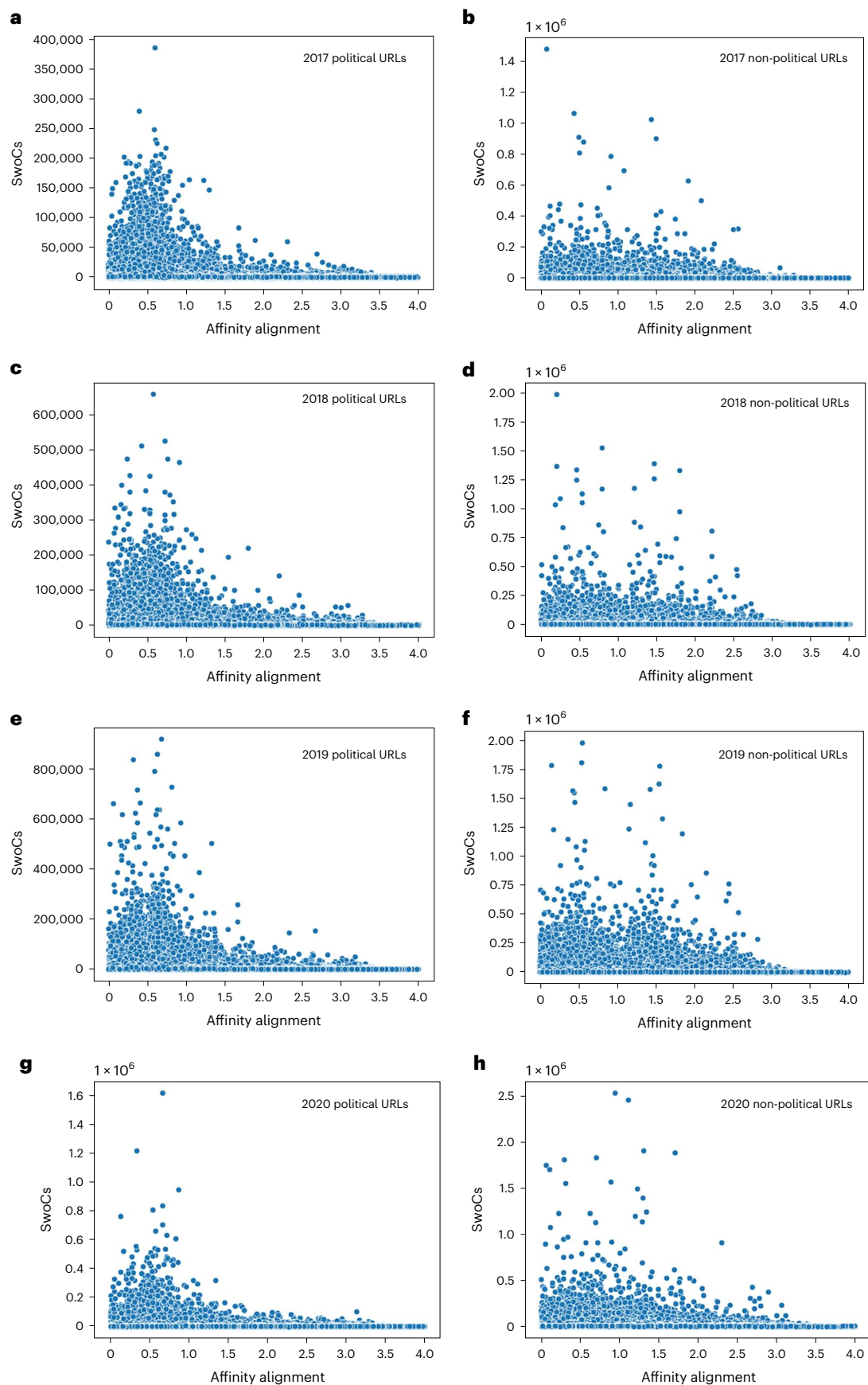


Fig. 3 | Relationship between affinity alignment and SwoCs of political and non-political URLs. **a**, 2017 May–November, political URLs: affinity alignment ($b = -994.88$, s.e. 4.59, 95% confidence interval (CI) $-1,003.88$ to -985.89 , $P < 0.001$), $R^2 = 0.054$. **b**, 2017 May–November, non-political URLs: alignment ($b = -431.19$, s.e. 1.99, 95% CI -435.09 to -427.30 , $P < 0.001$), $R^2 = 0.012$. **c**, 2018 midterm, political URLs: affinity alignment ($b = -983.45$, s.e. 5.81, 95% CI -994.83 to -972.08 , $P < 0.001$), $R^2 = 0.029$. **d**, 2018 midterm, non-political URLs: alignment ($b = -373.19$, s.e. 2.45, 95% CI -377.99 to -368.39 , $P < 0.001$), $R^2 = 0.005$.

e, 2019 May–November, political URLs: affinity alignment ($b = -880.65$, s.e. 6.12, 95% CI -892.65 to -868.65 , $P < 0.001$), $R^2 = 0.016$. **f**, 2019 May–November, non-political URLs: alignment ($b = -461.38$, s.e. 3.19, 95% CI -467.64 to -455.12 , $P < 0.001$), $R^2 = 0.004$. **g**, 2020 presidential, political URLs: affinity alignment ($b = -964.95$, s.e. 4.95, 95% CI -974.66 to -955.24 , $P < 0.001$), $R^2 = 0.021$. **h**, 2020 presidential, non-political URLs: alignment ($b = -408.23$, s.e. 2.33, 95% CI -412.80 to -403.66 , $P < 0.001$), $R^2 = 0.005$.

Table 2 | SwoCs and shares by five political user affinity groups

User affinity	SwoCs (% of total SwoCs)	Shares (% of total shares)	Percentage of SwoCs among shares
-2 = very liberal	7,940,478,846 (18.63%)	10,649,441,820 (18.89%)	74.56%
-1 = moderately liberal	9,405,668,831 (22.07%)	12,442,468,182 (22.07%)	75.59%
0 = neutral	7,782,109,480 (18.26%)	10,103,586,234 (17.92%)	77.02%
1 = moderately conservative	8,459,746,717 (19.85%)	11,172,584,511 (19.82%)	75.72%
2 = very conservative	9,028,481,339 (21.19%)	12,001,099,590 (21.29%)	75.23%
Total	42,616,485,213 (100%)	56,369,180,337 (100%)	75.60%

Aggregates from each user affinity group derives from 4,617 domains included in our analysis.

shares) came from moderately liberal users and very conservative users (Table 2). The total number of 42.6 billion shares without being clicked from those 4,617 top domains accounted for 75.6%, out of the 56.4 billion shares recorded. Across different political user affinity groups, the general proportion of shares that were not clicked amounted to ~75–77% of total shares (Table 2). This pattern was found even when we included lesser shared domains and all the months. Overall, SwoCs (80,845,962,779–80,853,679,349) account for 76.71% of shares (105,387,911,980–105,398,715,178) among all domains ($N = 334,200$) and URLs ($N = 38,750,429,130$) across 4 years, from 1 January 2017 to 31 December 2020. At the URL level, politically leaning (versus neutral) groups tended to share as well as share more without clicking, with more of those behaviours occurring among moderate to highly liberal groups in 2017 (Fig. 4a) and 2018 (Fig. 4c) for political URLs. This concentration started to shift to highly conservative groups in 2019 (Fig. 4e), and became stronger in 2020 (Fig. 4g). For non-political URLs, the amount of shares and SwoCs was smaller in size and also relatively even across different political affinity groups (Fig. 4b,d,f,h).

Is the SwoC phenomenon more apparent with false news?

Unlike domain-level data, URL-level data allow us to examine the effects of political affinity of users and content specifically tied to misinformation. When we examined URLs that were identified by third-party fact-checkers as false, we found that the absolute number of SwoCs was higher in conservative than in liberal political affinity groups. For instance, for the 2,969 URLs identified as false, approximately 17.5 million SwoCs came from the very conservative group, 14 million from the moderately conservative group, 3.6 million from the neutral group, 3.2 million from the moderately liberal group and 2.6 million from the very liberal group. Due to ‘differential privacy’ noise embedded in the dataset, the SwoC count for each political affinity group is identifiable in range rather than absolute values (see Supplementary Data Table 6 for the SwoC value ranges). For URLs classified as true, the absolute number of shares (including those without clicks) from liberals was higher than conservatives, with both liberals and conservatives sharing more without clicking compared with neutral users (Supplementary Data Table 7). Extended Data Fig. 1 shows the portion of SwoCs among all shares, as well as the number of shares and SwoCs per URL. To disambiguate whether the higher spread of misinformation by conservatives is due to their greater tendency to share false news, or because the content flagged as misinformation is

associated more closely with content that is circulated more among conservatives, we again selected URLs coming from domains that had an external media bias rating. Our findings suggest that the URLs identified as false by fact-checkers were more likely to come from conservative domains, while true news tended to come from liberal sites, suggesting domain as a cause for greater spread of misinformation by right-leaning users of the platform (see ‘Analysis of fact-checked URLs and those with media bias ratings’ section in Methods for statistical details of the analysis with fact-checked URLs). This confirms other researchers’ findings²⁵ from the 2020 election cycle that more misinformation came from the ‘conservative corner’ of Facebook, and extends it by showing that this asymmetry existed in the preceding years. In sum, our data show that the proportion of SwoCs among all shares of fact-checked URLs is within the 68–78% range for all five political affinity user groups. Analyses suggest that differences among the affinity groups in the volume of sharing false information stem from differences in the number of false URLs originating from news domains frequented by those groups.

Discussion

By analysing news sharing behaviours on Facebook across 4 years (that is, the top 4,617 shared domains and 35 million URLs during politically relevant time periods between 2017 and 2020), we found that sharing without clicking is indeed a widely prevalent phenomenon, much more than previously documented in the context of mainstream news domains³. We found that SwoC constitutes more than 75% of shares in the Facebook news sharing environment, which is quite telling as well as alarming in that the virality of political content on social media appears to be driven by superficial processing of headlines and blurbs rather than systematic processing of the core of the content. This could potentially explain why it is so common for misinformation to spread so quickly via social media; the vast majority of news stories that are shared on these platforms are shared without deliberation, let alone confirmation, of the underlying facts.

News sharing is a way of externalizing information to others³⁴, signalling that the information provider is confident and welcoming of future discussion with others in the shared news environment³⁵. Research¹² has shown that news consumers on Facebook feel more involved with a story when they share it rather than simply reading it, and this is even more so when they share it along with a question directed at their friends. This ideal of political information exchange

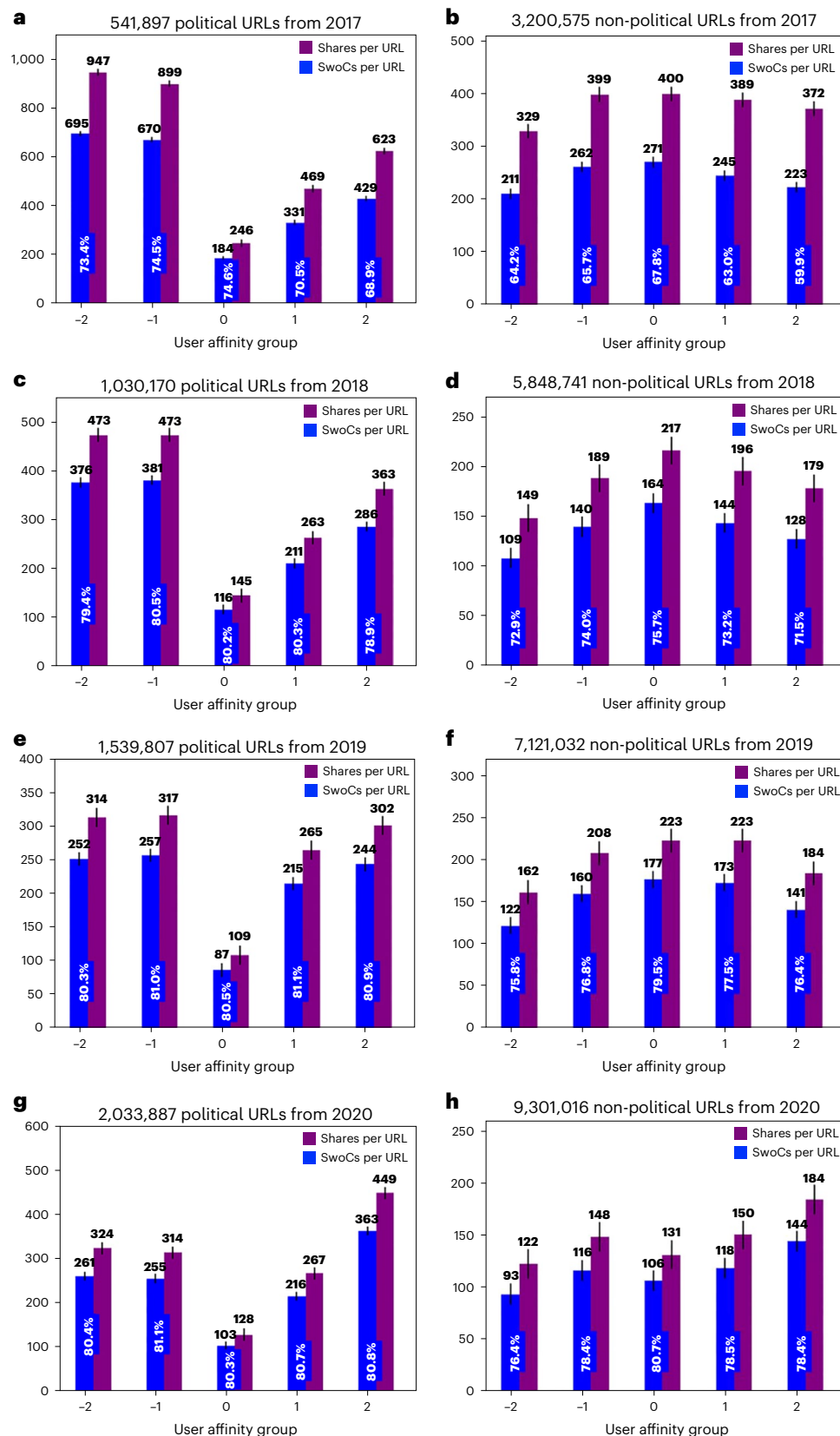


Fig. 4 | Shares and SwoCs of political and non-political URLs by five political user affinity groups. **a**, Political URLs from 2017. **b**, Non-political URLs from 2017. **c**, Political URLs from 2018. **d**, Non-political URLs from 2018. **e**, Political URLs from 2019. **f**, Non-political URLs from 2019. **g**, Political URLs from 2020. **h**, Non-political URLs from 2020. The data include error ticks on top of each bar that reflect the standard deviation of noise embedded in shares ($\sigma = 14$) and SwoCs ($\sigma = 10$). As such, since the errors are accounted for in the figures, we included all URLs (including

those with less than 2,395 shares). σ is a noise parameter calculated by Meta for each user action, including shares and SwoCs³¹. Due to the complexity of the data (in terms of the large URLs counts as well as ‘differential privacy’ noise embedded in aggregates for each URL, of which the noise only reduces as the aggregates are large enough), we summed up all the shares and SwoCs from each political user affinity group, divided them by the total URL counts from each year (as indicated in Table 1) and then calculated the ranges of shares and SwoC aggregates.

as part of a deliberative democracy can be seriously threatened by the widespread prevalence of SwoC discovered in our study. It prompts us to rethink the notion of news sharing and wonder about the depth of thought put into this highly prevalent action. It urges us to make a distinction between mindful news sharers who are truly engaged in the content and mindless news sharers who choose not to expend the cognitive effort to process the information before making their sharing decisions. While we could advise news consumers to distinguish between mindful and mindless news sharers, such vigilance is untenable given the aforementioned problem of information overload. Interface solutions that prevent sharers from sharing without reading first or at least nudge them towards more deliberation before sharing actions may be needed.

Different from shares with clicking, SwoCs can be far less deliberate and thoughtful, as users make the quick sharing decisions based not on the content of the news but on superficial cues either on the interfaces or in the headline. Both our domain- and URL-level analyses offer support to our prediction that politically extreme content was shared more without clicking. Our study also found that this behaviour is associated with partisan biases, with liberals (in 2017 through 2019) and, increasingly, conservatives (in 2020) more likely to perform this action than ideologically neutral users. While past research studying news sharing behaviours has found competing evidence in terms of whether users are more likely to share pro-attitudinal news articles³⁶ or articles that contain conflicting viewpoints³⁷, we found strong evidence that users are more likely to share ideologically consistent news content without clicking. This result is consistent with previous research showing that users rely on heuristic cues to make news sharing decisions on social media when facing information overload¹¹. In particular, we found that users are largely biased by their prior beliefs in making news sharing decisions, which is consistent with the notion of motivated reasoning³⁸.

Furthermore, given that the degree of affinity between users' and the content's political ideology positively predicts the volume of SwoCs, it appears that partisans are more eager to spread attitude-congruent political stories without bothering to vet them for factual accuracy, thus enhancing the probability of political misinformation enjoying widespread dissemination. These findings resonate with the notion that one's Facebook network tends to share substantially less news that is opposed to the group's general political ideology³⁹, which can be exacerbated by algorithmic suggestions reducing exposure to cross-cutting content⁴⁰ and thereby contributing further to ideological segregation by promoting exchanges within homogeneous networks⁴¹.

While our data suggest a strong pattern that partisans are more inclined to share news aligned with their ideology without reading it first, it is possible that this effect is driven by a small number of hyper-partisan users and communities within Facebook, as was found in Reddit, a platform where the majority of segregated news sharing occurs within extremely partisan segments and disproportionately more so in right-wing communities⁴². We also note that, by applying a minimum share threshold to select the domains included in our main analysis, we might have undercounted the SwoCs by conservatives because recent research by Muise et al.⁴³ indicates a greater number of less viewed (and, hence, less shared) right-leaning domains/sites than left-leaning ones. We would like to note that Muise et al.⁴³ applied the domain bias scores of Robertson et al.³³, which we also did in our study for supplementary analysis in 'Domain-level replication with external media bias ratings' section in Methods. When we visualized the distribution of the content affinity scores for the 4,617 domains, and compared it with the distributions of content affinity with URLs from the 2,636 domains for which we were able to apply the domain bias scores of Robertson et al.³³, both distributions suggest similar patterns (see Extended Data Fig. 2, with AllSides media bias ratings provided for comparison), without a significant loss of extreme right-leaning domains as demonstrated in Muise et al.⁴³. Future research is needed

to assess whether the sharing patterns discovered in this study are restricted to small pockets, or simply more pronounced in them, and the extent to which they influence larger societal phenomena such as political polarization and the epidemic of misinformation.

Our study's findings should be interpreted by keeping in mind certain caveats. The nature of the dataset precludes us from observing users' actual behaviours when browsing their social media feeds. In some circumstances, sharing after reading does not necessarily mean the users are more attentive to the information than when they share without reading. Likewise, not all SwoCs are necessarily mindless—some shares made without clicking the URL beforehand can still represent mindful resharing of familiar content that is perused by the user in another forum. Future lab observation studies could be conducted to uncover the nuances of these two types of sharing behaviour. In addition, with the ubiquitous nature of social media, users might be accessing them via different devices, which might affect their likelihood of engaging in SwoC. For instance, research has revealed that reading news on mobile phones could result in a spotlighting effect due to their smaller screens (compared with personal computers), which might contribute to more in-depth, instead of shallow, processing^{44,45}. Yet, other research suggests that users are usually less attentive when reading news or social media posts on their phones^{46,47} and are more likely to fall for misinformation on mobile phones than personal computers⁴⁸. More research is needed to understand differential rates of SwoC across different devices and platforms. This can have practical implications for curbing the spread of misinformation and promoting the quality of news and public affairs information on social media.

In conclusion, social media platforms would do well to learn from the behavioural patterns discovered in this study and accordingly design interfaces for promoting more systematic, rather than heuristic, processing of public affairs information appearing on them. One strategy is to embed general warnings (for example, letting users be aware of the prevalence of sharing without clicking), which can serve as a literacy message and promote non-specifically induced alertness about their news consumption⁴⁹. Another more effective strategy could be providing just-in-time alerts both to users who are about to share without clicking and to users who receive shared content (by informing them whether or not it has been clicked before sharing). This is akin to flagging a post as fake news, which has been shown to inhibit sharing of misinformation on Facebook^{50,51}. These strategies may serve to foster more informed and deliberate use of public affairs information appearing in our social media feeds.

Methods

While the dataset analysed in this study contains behavioural user data collected via the Facebook platform, it does not involve recruitment of, nor direct user responses from, human participants on an individual level. There is no personally identifiable information included in the dataset at the individual level, and our implementation of 'differential privacy' makes it "impossible to determine—in a way that is significantly better than random guessing—whether a specific user took an action in these data" (p. 9)³¹. In addition, we as researchers were not involved in the data collection process and were only allowed to access the dataset and performed analysis with a portion of the dataset relevant to our research questions of interest. Thus, this study does not constitute human participants research. The general data retrieval and analysis processes were approved by Pennsylvania State University's institutional review board under protocol ID STUDY00010564.

Source data characteristics

In February 2020, Social Science One in collaboration with Facebook made the largest social science dataset available to academic researchers, namely, 'Facebook URLs dataset'³¹. The dataset was designed to incorporate variables that can inform the effects of social media on elections and democracy. At the time of announcement, the dataset

contained Facebook users' behavioural data, such as likes and shares, pertaining to approximately 38 million URLs publicly shared over 100 times globally between January 2017 and July 2019. The dataset has been extended to incorporate more recent user behaviours over the years, with the most recent documentation of the dataset dated 12 April 2023 (version 10), upon which our results are based. This data version includes demographics of individuals who viewed, shared and otherwise interacted with web pages (URLs) shared on Facebook starting 1 January 2017 and up to and including 31 October 2022, which covers about 68 million URLs shared as an original post or reshared with 'public' privacy settings more than 100 times³¹. Considering our research goals focused on SwoCs, we extracted clicks, shares and SwoCs from 4,617 top-shared domains in the United States with political page affinity score assigned, stemming from all URLs shared on the platform during a 4-year period from January 2017 to December 2020. For the URL-level analysis, we extracted users' behavioural data (recorded in the United States, with political page affinity score assigned) amounting to 35 million URLs, across 4 years from 2017 to 2020, covering only May to November each year. All analyses reported in this article were confirmed and finalized by October 2024, after seeking and obtaining assurance from Meta that there have been no changes or updates made after October 2022 to the URL shares dataset, which is currently in maintenance mode.

The unit of data from the Facebook URLs dataset is based on URL-level information (that is, each data row represents information about each individual URL (not individual users)). One issue when dealing with this type of dataset is that a single or a particular group of URLs can reveal sensitive and latent personal attributes of users, such as their political affiliation and sexual orientation⁵². As a solution, 'differential privacy' prevents researchers from detecting whose information is included in the dataset by adding noise to the aggregated variables, so that those behavioural aggregates offer results within a statistically acceptable range instead of actual amount. While there are clear benefits of introducing such privacy-protective measures to user privacy and ethical research conduct, from a statistical standpoint, it may create measurement errors and statistical biases in analysing the dataset. To support precise reporting of the data patterns, Messing et al.³¹ provided noise parameters (σ) for the noise added for each user behavioural aggregate (for example, $\sigma = 40$ for clicks, $\sigma = 14$ for shares and $\sigma = 10$ for SwoCs) to allow data analysts to report statistically significant range for each aggregate while guaranteeing privacy protection to over 99% of users included in the dataset. To statistically test the relationships between variables, we ran ordinary least squares linear regression. To note, Meta and associated researchers offer a statistical package named Svinfer⁵³ to deal with datasets subject to 'differential privacy' wherein the noise parameters are known for predictor (independent) variables in linear and logistic regressions. However, in our case, we assume that the independent variables (that is, content affinity and affinity alignment) are not subject to noise for statistical testing because we applied Buntain et al.'s³² threshold. Hence, we set the noise parameter to zero in Svinfer. While our dependent variables (that is, shares and SwoCs) are not free from 'differential privacy' noise, Evans and King⁵⁴ note that "it creates only minor statistical problems" (p. 5). As can be expected with count data, we found our dependent variables to be heavily skewed, but opted to conduct linear regression instead of negative binomial regression. As Evans and King⁵⁴ note, "linear regression estimates give the best linear approximation to any functional form" regardless of the underlying distribution. We compared Svinfer's output with non-Svinfer ordinary least squares linear regression output (which we relied on for reporting the estimates required by this journal) and found that all coefficients match, with slight variations in standard errors that do not alter the statistical interpretation of our analyses.

Construction of major variables

While our main outcome variable, SwoC, is readily available in the dataset, the political valence of the URL is not. What the dataset offers

is a variable classifying political affinity of the users into five categories (which is one of our three predictors), only for the US data. The five categories (−2 to 2) "take the pages people follow and scale them into components associated with political affinity"³¹, based on Barberá et al.'s⁵⁵ approach that was applied to estimate ideological preferences of Twitter users. To compute Twitter users' ideological coordinates, Barberá et al.⁵⁵ identified 406 political Twitter accounts with relatively high ideological identification, such as all members of the Congress with more than 5,000 followers, the Democratic and Republican parties, and media outlets frequently classified as liberal (for example, @nytimes) or conservative (for example, @limbaugh). Then, they identified users with higher political interest (below the 20th and above the 80th percentile of ideology distribution among those who followed at least ten of those major political accounts) and added 800 political accounts most popular among those liberal and conservative users. In the end, 1,206 political accounts with over 3 million location-identified and/or active Twitter users following at least one of those accounts were obtained to classify each user's political affinity based on the political accounts they follow. This method was validated with external information, including a portion of those Twitter users that were matched with their voter registration history, statewide ideology averages and roll-call records of major political accounts. To calculate the political affinity of the content from the Facebook URLs dataset, Buntain et al.³² created a proxy by utilizing the political affinity of users. They aggregated the URL-level data into domain-level data and selected 2,629 domains, consisting of top 1% most shared US domains among the politically engaged Facebook users, to generate a political affinity estimate for each domain. Specifically, they divided shares for each domain based on the five political user affinity groups included in the Facebook dataset (−2 = very liberal, −1 = moderately liberal, 0 = neutral, 1 = moderately conservative, 2 = very conservative) and weighted those five affinity group indicators by shares documented for each group. They validated that, when the number of shares reaches a point where it is greater than the noise added, bias in the computed domain affinity becomes zero. Following this approach, we weighted five political group affinity values (−2 to 2) by the number of shares for each affinity group at both URL and domain levels to calculate the political affinity of content. For instance, if a URL/domain is shared 100 times by the −2 affinity group, 150 times by the −1 affinity group, 200 times by the 0 affinity group, 250 times by the 1 affinity group and 300 times by the 2 affinity group, the URL/domain score is calculated as $0.5 = [(100 \times -2) + (150 \times -1) + (200 \times 0) + (250 \times 1) + (300 \times 2)] / (100 + 150 + 200 + 250 + 300) = 500 / 1,000$. We apply shares as the main engagement metric to calculate the political affinity of content following prior applications to Facebook³⁹ and Twitter³³. Buntain et al.³² also found that the political affinity scores based on shares was not significantly different from those calculated on the basis of other engagement metrics such as views and likes. We note that this study only focused on sharing behaviours seen across URLs from users assigned with a specific political group affinity score (excluding behaviours witnessed from user groups unassigned with any political group affinity scores and, therefore, unaffected by the error in dataset reported by Meta in 2021, which accidentally excluded URL aggregates from US users with no political page affinity scores assigned). Affinity alignment between users and content on SwoC was calculated as the absolute value of the political affinity of content score subtracted by the user affinity group score. For instance, a URL/domain assigned a 0.5 score via the above calculation would have an affinity alignment score of 1.5 with the −1 affinity group ($= |0.5 - (-1)|$) and a score of 0.5 with the +1 affinity group ($= |0.5 - (+1)|$). This provided us a measure of the distance between the political stance of the content and that of the user. As associated user behaviours, we also compare the data patterns of SwoCs with overall shares (that is, number of users who shared the URL in a post or reshared such a post) and clicks (that is, number of users who clicked on the URL).

Data preparation process

For the domain-level analysis, we extracted 5,000 domains with the highest number of shared links in the United States from five political affinity groups (ranging from January 2017 to December 2020). However, we found that a small number of domains were missing information on some political affinity groups (for example, only having data from extremely conservative/liberal groups), which we filtered out, yielding a total of 4,975 top domains with scores from all five political user affinity groups. Even though top-shared domains tend to have sufficiently large shares to avoid bias created by added noise, Buntain et al.³² further identified a useful metric that determines the size of acceptable shares per domain to have statistical confidence over the calculated content affinity scores. They suggested that a minimum number of shares for each domain should be over 1.578 multiplied by $\sqrt{16 \times 5 \times 3 \times 7 \times \text{months of data coverage} \times \text{number of domain links} \times \text{noise estimate of shares}}$ for that domain within the Facebook dataset. Based on this equation, 357 domains did not exceed the required number of shares for their number of URL links and were, thus, excluded from our analysis. One outlier with extremely high shares and SwoCs (YouTube) was also excluded. In total, 4,617 domains were selected for our main analysis, including the following ten top-shared domains over the 4 years after youtube.com (6 billion shares): giphy.com (1,339 million shares), foxnews.com (1,022 million shares), cnn.com (1,012 million shares), nytimes.com (913 million shares), washingtonpost.com (855 million shares), tenor.co (834 million shares), nbcnews.com (651 million shares), dailywire.com (635 million shares), yahoo.com (578 million shares) and Breitbart.com (537 million shares).

The limitation of domain-level data is that they do not reflect content at the level of individual stories. Therefore, we also analysed URL-level data. Just as we replicated domain-level analysis with externally determined media bias ratings, we analysed sharing behaviours at the content level to offer more targeted insights relevant to the political context of this study. To do so, we analysed URL-level data across a 4-year span, for two politically relevant timeframes in 2018 and 2020 and two general timeframes for comparison in 2017 and 2019. In addition, each URL was classified as either political (1) or non-political (0) on the basis of a machine-learning-driven political classifier we developed for the purpose of this study, since the Facebook dataset does not offer any indicators of content type (that is, political versus non-political) at the URL level. Specifically, we created an indicator by using the Scikit-learn framework for unsupervised machine learning to perform topic modelling on the content of the URLs (for example, share_title and share_main_blurb) to compile an initial list of politically relevant keywords (that is, presidential, election, candidate, nominee, democrat, republican, independent, vote, Super Tuesday and congress), which were used to extract politically relevant URLs from the data. Next, we developed a supervised text classifier using the Scikit-learn framework. The classifier was trained using 2,000 manually labelled URLs from the data (1,000 politically relevant URLs based on keywords and 1,000 randomly selected URLs without any keyword restraints) for each year (amounting to a total of 8,000 URLs for 4 years) to balance political versus non-political URLs for manual coding and data training. We then tested a suite of machine learning modules and methods, including SVM, naive Bayes and logistic regression. A 10-fold cross validation approach was used to train and test the performance of different methods. We deployed random forest for the study, as it outperformed the other modules with performance indicators (that is, recall, precision and F1 score) showing above 90% for the 2018 dataset, which we used for data testing and training. Finally, the classifier was applied to the URLs based on the combination of share_title and share_main_blurb in each URL, where a value of 1 means the URL is politically relevant and a value of 0 means it is non-political. Similar to the domain-level analysis, we selected URLs that met the threshold of 2,395 shares for overcoming noise due to 'differential privacy', according to

Buntain et al.'s³² metric for individual URLs, for the May–November period (7 months) each year, from 2017 to 2020.

URL-level analyses of the associations between major variables (content affinity, alignment and SwoCs) and fake news

To test how content affinity of the political URLs predicts SwoCs, we only selected political URLs and excluded the URLs with assigned weighted affinity below −2 or above 2 for each year. We acknowledge that abnormal ranges in weighted affinity might have resulted from the insufficient number of shares for each individual URL due to the noise introduced by 'differential privacy'. Note that we did not need to exclude domains with weighted affinity below −2 and above 2 for the domain-level analysis, since the weighted affinity scores were distributed within the predicted range between −2 and 2. As noted earlier, we only selected URLs with shares over 2,395, following the suggestions of Buntain et al.³². In the 2017 dataset, 1,010,637 URLs (24.09%) out of 4,195,082 URLs met the share size requirement. In 2018, 1,178,093 URLs (16.67%) out of the 7,068,001 URLs met the share size requirement. In 2019, 1,658,672 URLs (15.79%) out of the 10,503,168 URLs met the share size requirement. In 2020, 2,017,701 URLs (14.95%) out of 13,494,326 URLs met the share size requirement. Note that, while the proportion of URLs that meet the size requirement across 4 years ranges from 15% to 24%, analyses with all URLs showed highly similar patterns (figures are identical to the human eye) with smaller coefficient values representing the effects of our major variables. In addition, one extreme outlier among the political URLs in 2017 was eliminated, since it was obscuring the visualization of the data patterns. The difference in data patterns between political and non-political URLs in 2017 data was found for 2018, 2019 and 2020 as well (Fig. 2). The coefficient and standard error values of content affinity on SwoCs for each year are reported in Supplementary Data Table 1.

To test how the alignment level between the URLs and the users affects SwoCs, we again excluded the URLs shared less than 2,395 times and with weighted affinity assigned below −2 or above 2. Data patterns distinguishing political and non-political URLs were similar for all four years (Fig. 3). The coefficient and standard error values of affinity alignment on SwoCs each year are reported in Supplementary Data Table 2.

Domain-level replication with external media bias ratings

As an alternative to the content affinity variable that we calculated on the basis of shares within the Facebook URLs dataset, we also conducted analyses with two types of domain affinity rating provided by an external organization. First, we obtained media bias ratings of 1,888 domains focused on news outlets offered by AllSides.com. Among the 4,617 domains included in our main analysis above, 817 domains had media bias ratings assigned by AllSides.com (−2 = left, −1 = lean left, 0 = neutral, 1 = lean right, 2 = right). Second, we borrowed media bias ratings developed by Robertson et al.³³, which has a relatively large sample of 19,022 domains, of which 2,636 domains were included in our dataset (−1 = very liberal to 1 = very conservative). When the content affinity variable was replaced with the media bias ratings from AllSides.com, there appeared a general surge in SwoCs for liberal domains and extremely conservative domains (Extended Data Fig. 3, left). The curvilinear pattern of content affinity on SwoC approached significance for the AllSides ratings, that is, the square of media bias scores ($b = 2,955,386.05$, s.e. 1,520,359.05, $P = 0.052$; regression model $R^2 = 0.006$) demonstrated a U-shaped curvilinear pattern as expected. However, the square of Robertson et al.'s domain ratings failed to show a statistically significant effect on SwoC ($b = -2,492,932.97$, s.e. 4,495,650.67, $P = 0.58$; regression model $R^2 = 0.0001$).

Next, we calculated the affinity alignment scores using the media bias ratings instead of share-based weighted content affinity. Note that Robertson et al.'s³³ bias ratings were multiplied by 2 when we calculated the affinity alignment score to match the 5-point scale to the 5-point user affinity group scores. A similar pattern emerged

showing that the more aligned the domain is to the user, the more the content from that domain was shared without clicking for both media bias scores (Extended Data Fig. 4). Statistically, the effect of affinity alignment on SwoC for both AllSides ($b = -1,861,031.19$, s.e. 246,745.23, $P < 0.001$; regression model $R^2 = 0.014$) and Robertson et al.'s³³ domain scores ($b = -936,668.26$, s.e. 148,771.50, $P < 0.001$; regression model $R^2 = 0.003$) was significant.

Analysis of fact-checked URLs and those with media bias ratings

To perform analysis on false news, we extracted all URLs (1) assigned with a specific rating from Third-Party Fact-Checking (TPFC) source, (2) shared on Facebook by users assigned a political affinity score (that is, political user affinity) and (c) within the period between 2017 and 2020 ($N = 25,007$). Category values include true, false, prank generator, false headline or mixture, opinion, satire, not eligible and not rated³¹. The above political classifier applied to each year was utilized to extract political content from the corpus of TPFC-rated URLs. Among all the fact-checked URLs ($N = 25,007$) coming from user groups with political affinity assigned over the span of 4 years, the majority had some misinformation, including those fact-checked as false ($N = 11,152$, 44.60%) and a mixture or false headline ($N = 1,619$, 6.47%). Unrated URLs occupied the next largest portion of the TPFC-rated URLs ($N = 8,685$, 34.73%), followed by those fact-checked as true ($N = 1,413$, 5.65%) (Supplementary Data Table 3). When we applied a machine learning-based political classifier for each year to select political (versus non-political) content (Supplementary Data Table 4), 7,560 (30.23%) were classified as political URLs. While the majority were fact-checked as false, the number of SwoCs per URL was higher for content identified as true (Supplementary Data Table 5; values in the brackets represent the upper and lower limit of the confidence intervals, accounting for embedded noise of respective user behaviours). The aggregates for each user affinity group are outlined in Supplementary Data Table 6 for false URLs and in Supplementary Data Table 7 for true URLs.

Among 7,560 political URLs with TPFC ratings, 1,828 URLs had media bias ratings from AllSides.com, and 4,076 from Robertson et al.³³ (Supplementary Data Table 8). We note that the number of URLs fact-checked as true is higher than those fact-checked as false, since the domains that have media bias scores are likely to include prestigious or popular news domains. First, for AllSides media bias ratings, over 70% (110 out of 150) of the URLs identified as false came from strongly conservative domains (Supplementary Data Table 9). URLs fact-checked as true were relatively balanced in the distribution of media bias scores, with more stemming from moderately (61 out of 189; 32.28%) to strongly (47 out of 189; 24.87%) liberal sources (Supplementary Data Table 10). While the sheer number of shares and SwoCs is larger for domains with higher URL counts, when we examined the user data per individual link, higher shares and SwoCs of both true and false news were witnessed among more liberal sites than conservative ones (Extended Data Fig. 5). Second, Robertson et al.'s³³ media bias ratings were calculated for each domain. The majority of false news shared without being clicked came from moderately (359 out of 966; 37.16%) to highly conservative (434 out of 966; 44.93%) domains. These domains accounted for nearly 70% of SwoCs of false news (Supplementary Data Table 11), whereas they attracted half that proportion (35%) of SwoCs of true news (Supplementary Data Table 12).

Reporting summary

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

Data availability

Access to Facebook data was granted via Social Science One, with all researchers in need of access to Facebook data signing a Research

Data Agreement between Facebook, Inc. and the Grantee institution (Pennsylvania State University). Details about these data, including the codebook, can be found at the Social Science One Dataverse repository at <https://socialscience.one/facebook-dataverse>.

Code availability

Researchers were allowed special access to a JupyterHub system that provides an interface to query the data. Information to obtain authentication and access to JupyterHub, as well as resources for data analysis, was communicated through an invite-only GitHub organization (<https://github.com/fb-researchtool-ssl>). All data were retrieved, prepared and analysed via Python programming language (version 3.10.11) with the following packages: NLTK (version 3.8.1) and scikit-learn (version 1.2.2). Examples of major Python codes used for data retrieval, preparation and analysis can be found via GitHub at <https://github.com/geocomplexity/SwoCMetaURL>.

References

- Chen, S. & Chaiken, S. in *Dual-Process Theories in Social Psychology* (eds Chaiken, S. & Trope, Y.) 73–96 (The Guilford Press, 1999).
- Gabelkov, M., Ramachandran, A., Chaintreau, A. & Legout, A. Social clicks: what and who gets read on Twitter? In *Proc. 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science* (eds Eager, D. & Williamson, C.) 179–192 (Association for Computing Machinery, 2016).
- Holmström, J. et al. Do we read what we share? Analyzing the click dynamic of news articles shared on Twitter. In *Proc. 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (eds Spezzano, F. et al.) 420–425 (Association for Computing Machinery, 2019).
- Schäfer, S. Illusion of knowledge through Facebook news? Effects of snack news in a news feed on perceived knowledge, attitude strength, and willingness for discussions. *Comput. Hum. Behav.* **103**, 1–12 (2020).
- Ohme, J. & Mothes, C. News snacking and political learning: changing opportunity structures of digital platform news use and political knowledge. *J. Inf. Technol. Polit.* 1–15 (2023).
- Andrew, B. C. Media-generated shortcuts: do newspaper headlines present another roadblock for low-information rationality? *Harv. Int. J. Press Polit.* **12**, 24–43 (2007).
- Lee, N. Y. Headlines for summarizing news or attracting readers' attention? Comparing news headlines in South Korean newspapers with *The New York Times*. *Journalism* **23**, 892–909 (2022).
- Carcioppolo, N., Lun, D. & McFarlane, S. J. Exaggerated and questioning clickbait headlines and their influence on media learning. *J. Media Psychol.* **34**, 30–41 (2021).
- Potthast, M. et al. Clickbait detection. In *Advances in Information Retrieval. ECIR 2016. Lecture Notes in Computer Science* Vol. 9626 (eds Ferro, N., et al.) 810–817 (Springer International Publishing, 2016).
- Potthast, M. et al. Crowdsourcing a large corpus of clickbait on Twitter. In *Proc. 27th International Conference on Computational Linguistics* (eds Bender, E. et al.) 1498–1507 (Association for Computational Linguistics, 2018).
- Schmitt, J. B., Debbelt, C. A. & Schneider, F. M. Too much information? Predictors of information overload in the context of online news exposure. *Inf. Commun. Soc.* **21**, 1151–1167 (2018).
- Oeldorf-Hirsch, A. & Sundar, S. S. Posting, commenting, and tagging: effects of sharing news stories on Facebook. *Comput. Hum. Behav.* **44**, 240–249 (2015).
- Lee, J. Y. & Sundar, S. S. To Tweet or to Retweet? That is the question for health professionals on Twitter. *Health Commun.* **28**, 509–524 (2013).

14. Wang, S., Chu, T. H. & Huang, G. Do bandwagon cues affect credibility perceptions? A meta-analysis of the experimental evidence. *Commun. Res.* **50**, 720–744 (2023).
15. Fu, W. W. & Sim, C. C. Aggregate bandwagon effect on online videos' viewership: value uncertainty, popularity cues, and heuristics. *JASIST* **62**, 2382–2395 (2011).
16. Kim, J. W. They liked and shared: effects of social media virality metrics on perceptions of message influence and behavioral intentions. *Comput. Hum. Behav.* **84**, 153–161 (2018).
17. Mourão, R. R. & Robertson, C. T. Fake news as discursive integration: an analysis of sites that publish false, misleading, hyperpartisan and sensational information. *J. Stud.* **20**, 2077–2095 (2019).
18. Molina, M. D. et al. Does clickbait actually attract more clicks? Three clickbait studies you must read. In *Proc. 2021 CHI Conference on Human Factors in Computing Systems* (eds Barbosa, S. et al.) 1–19 (Association for Computing Machinery, 2021); <https://doi.org/10.1145/3411764.3445753>
19. Sude, D. & Knobloch-Westerwick, S. In *Knowledge Resistance in High-Choice Information Environments* (eds Strömbäck, J. et al.) 88–105 (Routledge, 2022).
20. Knobloch-Westerwick, S., Mothes, C. & Polavin, N. Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Commun. Res.* **47**, 104–124 (2020).
21. Ling, R. Confirmation bias in the era of mobile news consumption: the social and psychological dimensions. *Digit. J.* **8**, 596–604 (2020).
22. Ross, R. M., Rand, D. G. & Pennycook, G. Beyond 'fake news': analytic thinking and the detection of false and hyperpartisan news headlines. *Judgm. Decis. Mak.* **16**, 484–504 (2021).
23. Messing, S. & Westwood, S. J. Selective exposure in the age of social media: endorsements trump partisan source affiliation when selecting news online. *Commun. Res.* **41**, 1042–1063 (2014).
24. Ziemke, D. A. Selective exposure in a presidential campaign contingent on certainty and salience. *Ann. Int. Commun. Assoc.* **4**, 497–511 (1980).
25. González-Bailón, S. et al. Asymmetric ideological segregation in exposure to political news on Facebook. *Science* **381**, 392–398 (2023).
26. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. Fake news on Twitter during the 2016 US presidential election. *Science* **363**, 374–378 (2019).
27. Iyengar, S. & Hahn, K. S. Red media, blue media: evidence of ideological selectivity in media use. *J. Commun.* **59**, 19–39 (2009).
28. Hoffman, M. L. Some psychodynamic factors in compulsive conformity. *J. Abnorm. Soc. Psychol.* **48**, 383–393 (1953).
29. Tarr, H. & Lorr, M. A comparison of right-wing authoritarianism, conformity and conservatism. *Personal. Individ. Differ.* **12**, 307–311 (1991).
30. Molina, M. D., Wang, J., Sundar, S. S., Le, T. & DiRusso, C. Reading, commenting and sharing of fake news: how online bandwagons and bots dictate user engagement. *Commun. Res.* **50**, 667–694 (2023).
31. Messing, S. et al. Facebook privacy-protected full URLs data set [V10]. *Harvard Dataverse* (2020).
32. Buntain, C., Bonneau, R., Nagler, J. & Tucker, J. A. Measuring the ideology of audiences for web links and domains using differentially private engagement data. *Proc. Int. AAAI Conf. Web Soc. Media* **17**, 72–83 (2023).
33. Robertson, R. E. et al. Auditing partisan audience bias within Google search. *Proc. ACM Hum.-Computer Interact.* **2**, 1–22 (2018).
34. Hendriks, P. Why share knowledge? The influence of ICT on the motivation for knowledge sharing. *Knowl. Process Manag.* **6**, 91–100 (1999).
35. van den Hooff, B. & de Leeuw van Weenen, F. Committed to share: commitment and CMC use as antecedents of knowledge sharing. *Knowl. Process Manag.* **11**, 13–24 (2004).
36. An, J., Quercia, D. & Crowcroft, J. Fragmented social media: a look into selective exposure to political news. In *Proc. 22nd International Conference on World Wide Web* 51–52 (Association for Computing Machinery, 2013); <https://doi.org/10.1145/2487788.2487807>
37. Morgan, J. S., Lampe, C. & Shafiq, M. Z. Is news sharing on Twitter ideologically biased? In *Proc. 2013 Conference on Computer Supported Cooperative Work* 887–896 (Association for Computing Machinery, 2013).
38. Taber, C. S. & Lodge, M. Motivated skepticism in the evaluation of political beliefs. *Am. J. Political Sci.* **50**, 755–769 (2006).
39. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
40. Guess, A. M. et al. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* **381**, 398–404 (2023).
41. Stroud, N. J. Polarization and partisan selective exposure. *J. Commun.* **60**, 556–576 (2010).
42. Mok, L., Inzlicht, M. & Anderson, A. Echo tunnels: polarized news sharing online runs narrow but deep. *Proc. Int. AAAI Conf. Web Soc. Media* **17**, 662–673 (2023).
43. Muise, D. et al. Quantifying partisan news diets in Web and TV audiences. *Sci. Adv.* **8**, eabn0083 (2022).
44. Kim, K. J. & Sundar, S. S. Mobile persuasion: can screen size and presentation mode make a difference to trust? *Hum. Commun. Res.* **42**, 45–70 (2016).
45. Alasmari, T. The effect of screen size on students' cognitive load in mobile learning. *J. Educ. Teach. Learn.* **5**, 280–295 (2020).
46. Dunaway, J., Searles, K., Sui, M. & Paul, N. News attention in a mobile era. *J. Comput. Mediat. Commun.* **23**, 107–124 (2018).
47. Keib, K. et al. Living at the speed of mobile: how users evaluate social media news posts on smartphones. *Commun. Res.* **49**, 1016–1032 (2021).
48. Liao, M., Wang, J., Chen, C. & Sundar, S. S. Less vigilant in the mobile era? A comparison of information processing on mobile phones and personal computers. *N. Media Soc.* <https://doi.org/10.1177/14614448231209475> (2023).
49. Ecker, U. K., Lewandowsky, S. & Tang, D. T. Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Mem. Cogn.* **38**, 1087–1100 (2010).
50. Mena, P. Cleaning up social media: the effect of warning labels on likelihood of sharing false news on Facebook. *Policy Internet* **12**, 165–183 (2020).
51. Nekmat, E. Nudge effect of fact-check alerts: source influence and media skepticism on sharing of news misinformation in social media. *Soc. Media Soc.* **6**, 2056305119897322 (2020).
52. Kifer, D., Messing, S., Roth, A., Thakurta, A. & Zhang, D. Guidelines for implementing and auditing differentially private systems. Preprint at <https://arxiv.org/abs/2002.04049v2> (2020).
53. Facebook Open Research and Transparency. GitHub – facebookresearch/svinfer. [GitHub https://github.com/facebookresearch/svinfer](https://github.com/facebookresearch/svinfer) (2020).
54. Evans, G. & King, G. Statistically valid inferences from differentially private data releases, with application to the Facebook URLs dataset. *Polit. Anal.* **31**, 1–21 (2023).
55. Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A. & Bonneau, R. Tweeting from left to right: Is online political communication more than an echo chamber? *Psych. Sci.* **26**, 1531–1542 (2015).

Acknowledgements

This research was supported by Social Media and Democracy Research Grants, awarded by the Social Science Research Council, with access to Facebook data via Social Science One (grantees: S.S.S., G.C., J.W. and J.Y.). The grants are funded by a group of charitable foundations including the Laura and John Arnold Foundation, the Children's Investment Fund Foundation, the Democracy Fund, the William and Flora Hewlett Foundation, the John S. and James L. Knight Foundation, the Charles Koch Foundation, Omidyar Network and the Alfred P. Sloan Foundation. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. See more information about the grant project at <https://www.ssrc.org/grantees/are-we-more-willing-to-speak-up-when-we-share-without-clicking-a-study-of-political-content-sharing-on-facebook/>.

Author contributions

Conceptualization: S.S.S., E.C.S., M.L., J.Y., J.W. and G.C. Methodology: S.S.S., E.C.S., M.L., J.Y., J.W. and G.C. Investigation: S.S.S., E.C.S., M.L., J.Y., J.W. and G.C. Data wrangling: E.C.S., J.Y. and M.L. Visualization: E.C.S., M.L., S.S.S. and J.Y. Funding acquisition: S.S.S., J.Y., J.W. and G.C. Project administration: S.S.S. and E.C.S. Supervision: S.S.S. Writing—original draft: S.S.S., E.C.S., M.L. and J.Y. Writing—review and editing: S.S.S., E.C.S., M.L., J.Y., J.W. and G.C.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41562-024-02067-4>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41562-024-02067-4>.

Correspondence and requests for materials should be addressed to S. Shyam Sundar.

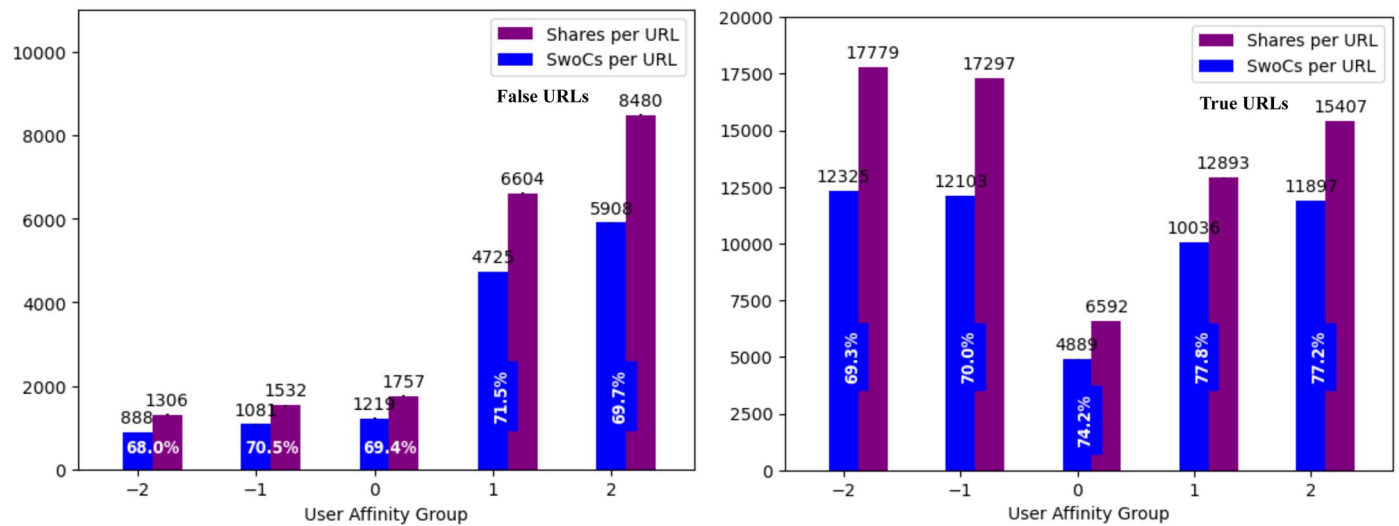
Peer review information *Nature Human Behaviour* thanks Porismita Borah, Daniel Muise and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

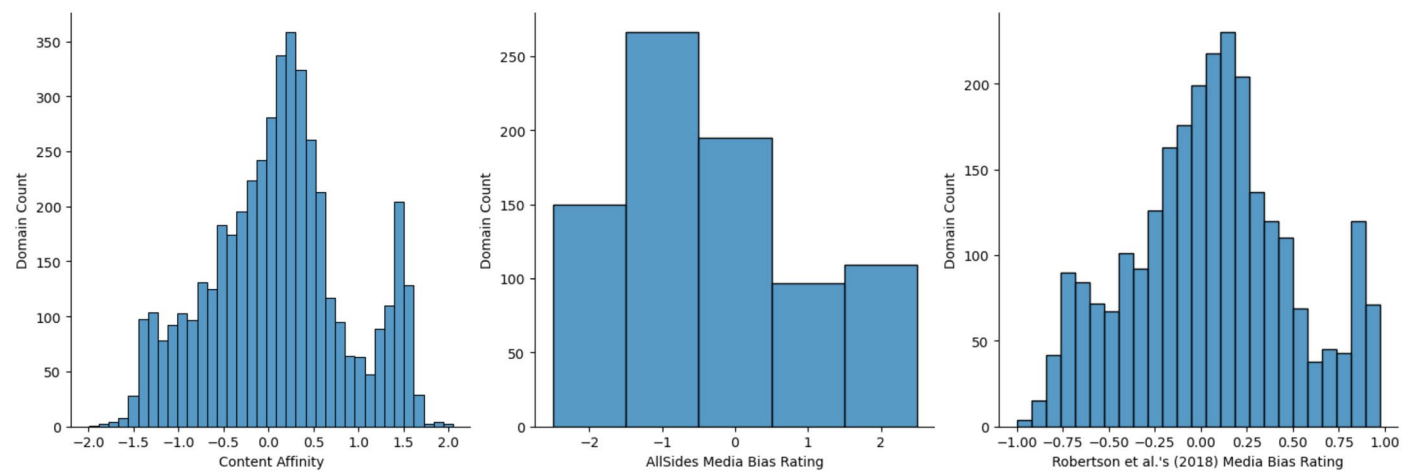
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

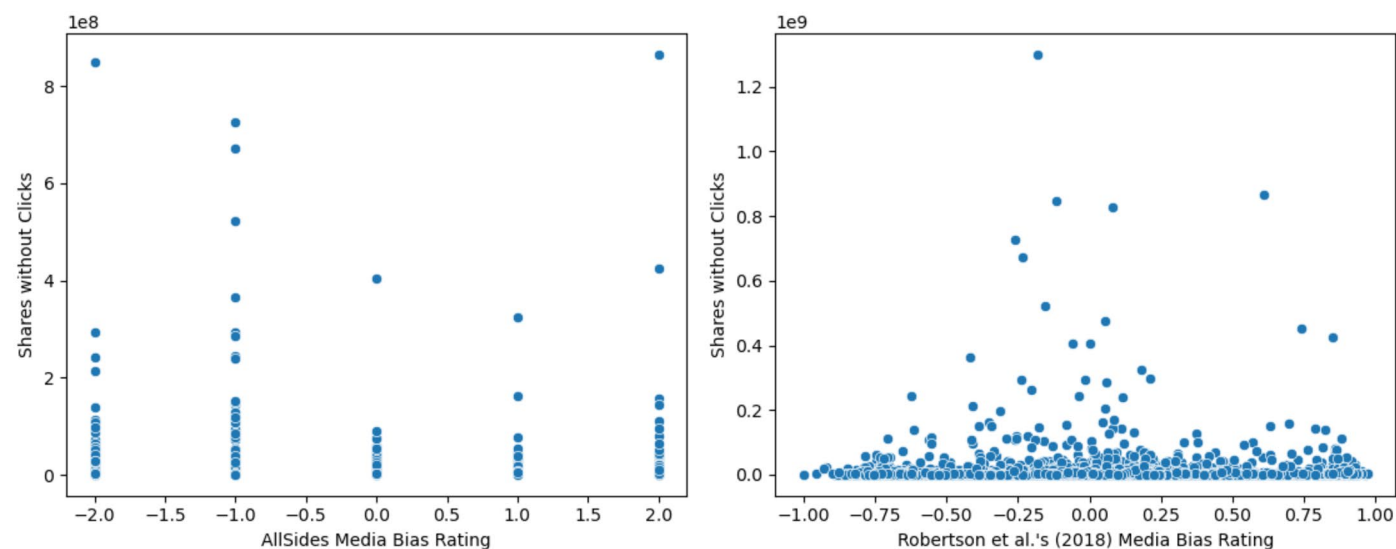
© The Author(s), under exclusive licence to Springer Nature Limited 2024



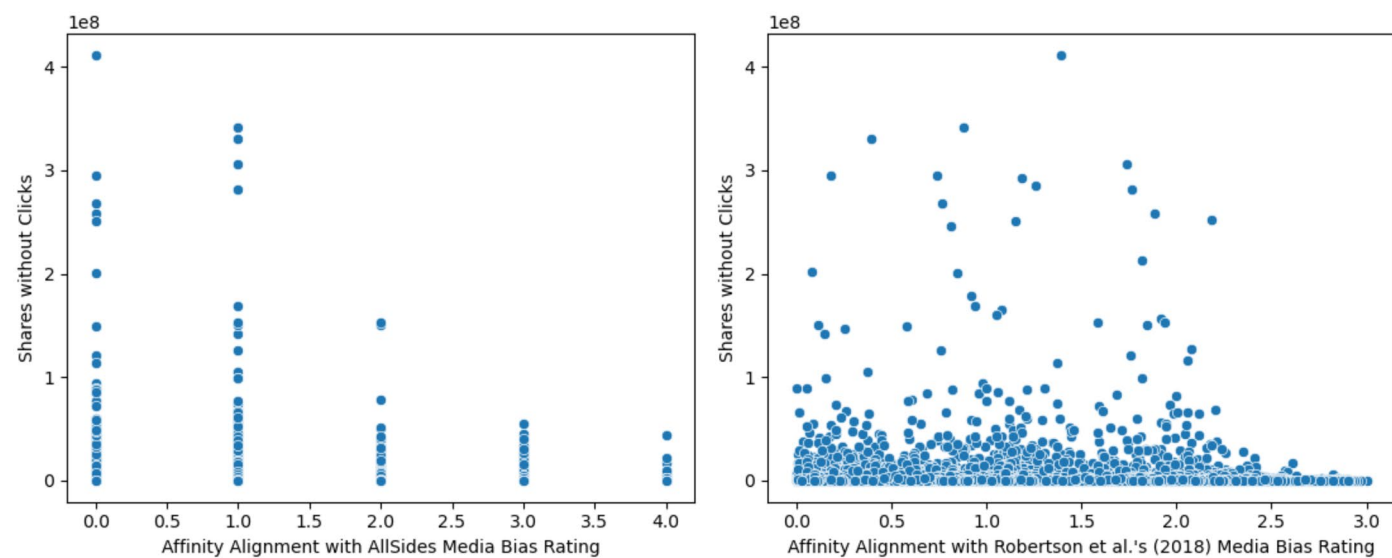
Extended Data Fig. 1 | SwoCs and shares of political URLs rated as false (left) and true (right) news by five political affinity groups. While hardly visible, error ticks are applied to account for noise embedded to shares ($\sigma = 14$) and shares without clicks ($\sigma = 10$). σ is a noise parameter calculated by Meta for each user action, including shares and shares without clicks.



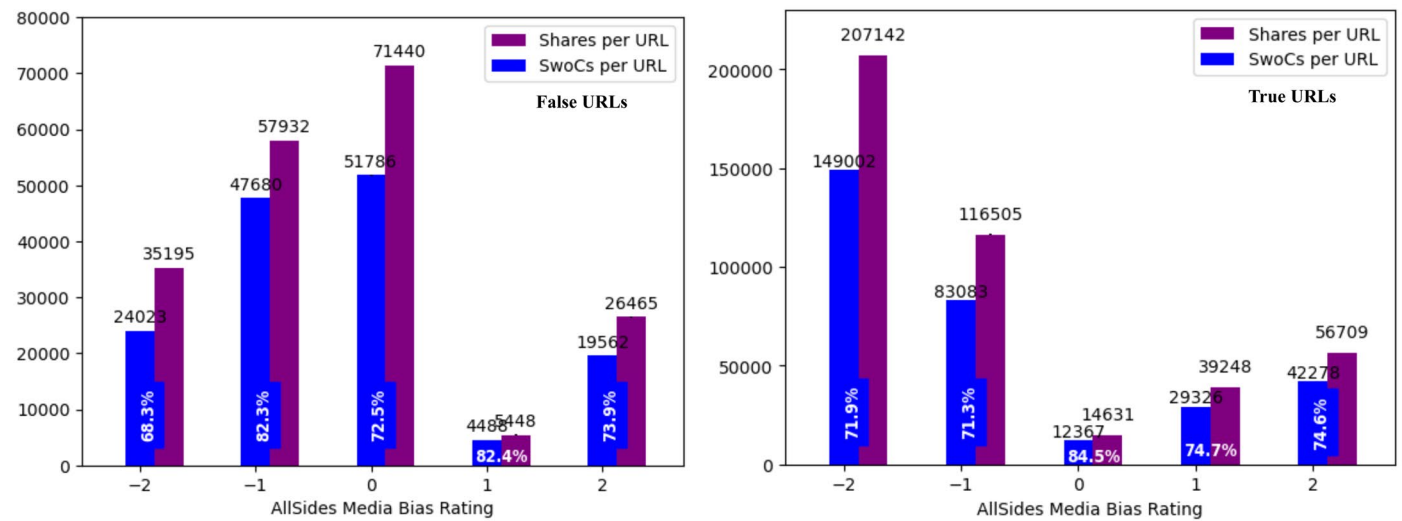
Extended Data Fig. 2 | Distribution of domain-level content affinity. Distribution of all 4,617 domains (left), 817 domains with AllSides media bias ratings assigned (middle), and 2,636 domains with Robertson et al.'s (33) media scores assigned (right).



Extended Data Fig. 3 | Relationship between domain-level content affinity and SwoCs among domains with external ratings of media bias. Data based on AllSides (left) and Robertson et al.'s (33) (right) ratings.



Extended Data Fig. 4 | Relationship between domain-level affinity alignment and SwoCs among domains with external ratings of media bias. Data based on AllSides (left) and Robertson et al.'s (33) (right) ratings.



Extended Data Fig. 5 | SwoCs and shares of fact-checked political URLs from media sources rated by AllSides. Data for false (left) and true (right).

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 (<i>n</i>) 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
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	Researchers analyzed a social science dataset available to academic researchers, labeled "Facebook URLs dataset", created and granted access by Social Science One in collaboration with Facebook (Meta). Researchers were not involved in the data collection process, and were only allowed to access the dataset and performed analysis with a portion of the dataset relevant to their research questions of interest. Thus, no software was used for data collection, but we used Python (version 3.10.11) for data retrieval.
Data analysis	As grantees with an accepted proposal, we were allowed special access to a JupyterHub system that provides an interface to query the data. Information to obtain authentication and access to JupyterHub, as well as resources for data analysis, were communicated through an invite-only GitHub organization (https://github.com/fb-researchtool-ss1). All data were retrieved, prepared, and analyzed using Python (version 3.10.11) with the following packages: NLTK (version 3.8.1) and scikit-learn (version 1.2.2). Example of major Python codes used for data retrieval, preparation, and analysis are included in supplementary materials as readme files.

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](#)

Access to Facebook data was granted via Social Science One, with all researchers in need of access to Facebook data signing a Research Data Agreement between Facebook, Inc. and the Grantee institution (Pennsylvania State University). Details about these data, including the codebook, can be found at the Social Science One Dataverse repository: <https://socialscience.one/facebook-dataverse>. Detailed information about the codes we used for data retrieval are included the Code Availability statement in our manuscript.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	The study does not include any findings or analysis related to sex and gender.
Reporting on race, ethnicity, or other socially relevant groupings	The only socially relevant categorization variable used in this study is the political affinity of users. However, the unit of data was based on the social media content (e.g., URLs or domains) and available in aggregate form, not at the level of individual users. In other words, the size of political affinity group was categorized for each content. For instance, the number of shares for each URL was classifiable based on 5 different political user groups (ranging from liberal = -2 to conservative = +2), which was calculated and readily available by the dataset offered by Social Science One in collaboration with Facebook.
Population characteristics	See above
Recruitment	There is no recruitment of human participants involved in this study.
Ethics oversight	This research was approved by Pennsylvania State University's Institutional Review Board (IRB).

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

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	The study conducted quantitative analysis of "shares without clicks" for all URLs publicly shared on Facebook between 2017 and 2020.
Research sample	Our unit of analysis is URL, not individuals. The study analyzed 4,523 top-shared domains aggregated from the URL-level data across the 4-year period. URL data from a politically relevant timeframe in 2018 from May to November (i.e., the 2018 midterm election season) as well as in 2020 from May to November (i.e., the 2020 presidential election season) were also analyzed. To offer comparison points, data from May to November for 2017 and 2019 were added. As a result, the study extracted users' behavioral data pertaining to about 35 million URLs over four years (4,195,082 URLs from 2017, 7,068,001 URLs from 2018, 10,503,165 URLs from 2019, and 13,480,583 URLs from 2020). Since this dataset is not structured at the individual user-level (vs. URL or domain content-level), we cannot offer demographic information or distributions at the user level. As indicated in the manuscript, the dataset "includes demographics of individuals who viewed, shared, and otherwise interacted with web pages (URLs) shared on Facebook starting January 1, 2017, and up to and including October 31, 2022, which covers about 68 million URLs shared as an original post or reshared with "public" privacy settings more than 100 times." We selected certain time frames as explained above to extract URL-level data, appropriate to answer our research questions for different years.
Sampling strategy	The sampling strategy of the dataset we accessed for our study is to include all URLs shared on Facebook starting January 1, 2017, and up to and including October 31, 2022, which covers about 68 million URLs shared as an original post or reshared with "public" privacy settings more than 100 times. We were not involved in the data collection process. We selected and filtered out URLs, data points and variables relevant to their research questions, from the readily available data frames offered by Social Science One in collaboration with Facebook. Because of no association with data collection, we cannot confirm the data sampling procedure. In

	addition, no sample size calculation was performed, and sample sizes (of URLs or domains) were chosen based on political relevance (e.g., the 2018 midterm election season) or data size requirements to reduce noise (e.g., 4,523 top-shared domains).
Data collection	Researchers were not involved in the data collection process. Instead, they analyzed a portion of dataset, offered by Social Science One in collaboration with Facebook, associated with their research questions of interest. This study was not an experiment, thus blinding procedures are not applicable.
Timing	The dataset used by the researchers, namely Facebook URLs dataset, was released in February 2020, pertaining to approximately 38 million URLs publicly shared over 100 times globally between January 2017 and July 2019, and has been extended to incorporate more recent user behaviors over the years. This study's analysis covers data collected over 4 years, from 2017 to 2020. We started data extraction and analysis during September 2020 and as noted in our manuscript: "All analyses reported in this article were confirmed and finalized in September 2024, after seeking and obtaining assurance from Meta that there have been no changes or updates made after October 2022 to the URL Shares dataset, which is currently in maintenance mode."
Data exclusions	Domains or URLs that did not meet the size requirement of acceptable shares to have statistical confidence over the calculated content affinity scores were excluded. For major analyses, domains or URLs having weighted affinity below -2 and above 2 were also excluded. Other exceptional exclusions such as outliers (e.g., domain missing all political affinity group information, domain with extremely high shares without clicking) are also explained within the main manuscript.
Non-participation	This study does not involve participant recruitment. Data are based on social media content (e.g., URLs or domains), not individual users.
Randomization	This study does not follow an experimental design, thus does not require randomization of participant allocation to different conditions. Controls or covariates are not applicable to our analyses either because we are dealing with content data, not people data, in an effort to detect (count) the nature of content that is shared most often without clicking.

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		Methods	
n/a	Involved in the study	n/a	Involved in the study
☒	☐ Antibodies	☒	☐ ChIP-seq
☒	☐ Eukaryotic cell lines	☒	☐ Flow cytometry
☒	☐ Palaeontology and archaeology	☒	☐ MRI-based neuroimaging
☒	☐ Animals and other organisms		
☒	☐ Clinical data		
☒	☐ Dual use research of concern		
☒	☐ Plants		

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a