Web Search Personalization During the US 2020 Election

Ulrich Matter[1] and Roland Hodler[1]

[1]Department of Economics, University of St.Gallen

December 18, 2024

Online Appendix

Table of Contents

A Technical Implementation

Our study's technical realization is based on custom-made software designed to set up, configure, and control synthetic web users ("web bots") who emulate human user characteristics and device fingerprints, and human-like browsing and search behavior. Figure D2 in the Online Appendix illustrates the key components of the application architecture. To run the users' browsing sessions, we rely on a specialized remote web driver. The web browsing of our synthetic users is not recognizable as being run by browser automation (unlike browsers automated by a standard web driver such as Selenium). Automated browser instances are then linked to two components that help us emulate unique human user characteristics: a fingerprint manager (allowing the user to appear as if it uses a specific operating system and hardware setting), and a residential proxy service (allowing the user to access the Web through a residential IP address in a specific US city, which would not be possible with common proxy servers or VPN services). All web traffic issued and received by users is recorded. This allows us to see what a given user did at any point in time and what the user was exposed to on Google and other websites. This part of the Online Appendix complements the main text by discussing how we verify our implementation.

A1 *Verification of fingerprinting and geolocations*

We evaluate user appearance with a set of security testing tools provided by BrowserLeaks (`www.browserleaks.com`). To this end, we extract and parse the reports generated by BrowserLeaks for each of our synthetic users and verify a) whether the synthetic users' device and network connection are consistent with the intended configuration, and b) whether our users' fingerprints are, in fact, all unique (and hence, our users can be successfully tracked and uniquely identified based on them). Figure D3 shows a sample screenshot taken from such a verification test.

Second, we monitor the residential proxy servers via an independent real-time geolocation service at the beginning and at the end of each user's browsing and search sessions. Figure D4 shows the third-party verified geolocations. Blue crosses mark the verified coordinates of users when browsing and searching (all browsing sessions of all users are included in the plot), and orange circles highlight the official city coordinates to which synthetic users were assigned during the study. Synthetic users were generally recognized as being located in their assigned city, with very few exceptions. We see that very few of the thousands of geocoded browsing sessions (blue crosses) are outside of the orange circles.

B Input Data and Synthetic User Configuration

This part of the Online Appendix provides a detailed discussion of the configuration of the synthetic users' browsing and search behavior, with a focus on the data-driven selection of popular domains, partisan domains, non-partisan search terms, partisan search terms, and the experimental election-related search terms. It also explains how we validate our selection of partisan search terms.

B1 Popular domains

We compile the set of popular domains from a list of the most frequently visited domains provided by Ahrefs (2020) – a software company that specializes in providing tools and services for search engine optimization and compiles large-scale web traffic data on millions of websites. Based on these web traffic data, Ahrefs (2020) publishes yearly estimates of the most visited websites per country under `https://ahrefs.com/top` (previously under `https://ahrefs.com/blog/most-visited-websites`). We collected the list of the 100 most visited US domains from `https://ahrefs.com/blog/most-visited-websites` on October 1st, 2020 (reflecting the ranking of the most visited website in the US as of May 2020 according to Ahrefs, 2020). Table E2 lists these domains.

Every day, we randomly draw 1–3 of these 100 most popular domains for each user individually and let them visit these domains.

B2 Partisan domains

Based on a collection of over 140M partisan tweets issued during the 2018 mid-term elections Wrubel, Littman and Kerchner (2019), we identify the top 100 domains mostly referred to by Democratic supporters and the top 100 domains mostly referred to by Republican supporters. We do so by first extracting and parsing all URLs appearing in pro-Democrats tweets and in pro-Republican tweets. We filter out URLs containing the domain twitter.com, abbreviated URLs, and domains no longer in use by October 2020. Following the idea of detecting partisan phrases in text by Gentzkow and Shapiro (2010), we then count the number of times a domain was referred to in a tweet by Democrats or Republicans and compute the 'partisanship' of each domain i as follows. For each domain i, we compute

$$\chi_i^2 = \frac{(f_{ir} f_{\sim id} - f_{id} f_{\sim ir})^2}{(f_{ir} + f_{id})(f_{ir} + f_{\sim ir})(f_{ir} + f_{\sim ir})(f_{\sim ir} + f_{\sim id})}, \tag{B1}$$

where f_{id} (f_{ir}) denote the total number of times domain i is mentioned in a tweet by Democratic (Republican) supporters, and $f_{\sim id}$ ($f_{\sim ir}$) denote the total number of times a domain other than i is referred to in a tweet by Democratic (Republican) supporters. A higher χ_i^2 value indicates that i is primarily mentioned by supporters of one of the two parties. Given f_{id}, f_{ir} and χ_i^2, we can select the 100 most partisan domains used by Democratic supporters and the 100 most partisan domains used by Republican supporters.

We then randomly assign each Democratic (Republican) user 10 of the 100 most Democratic (Republican) domains. Tables E3 and E4 show the most partisan domains that were randomly assigned to our Democratic and Republican users, respectively (see Matter and Hodler (2024) for the corresponding dataset).

B3 Non-partisan search terms

Non-partisan search terms were created using a Python script designed to ensure a diverse range of broadly used, yet non-partisan queries.[1] The generation of terms is based on incorporating common terms, country-specific modifiers, unit conversions, and simple arithmetic equations:

1) **Common Search Terms (ahrefs.com)**: These are based on the top 100 search terms most commonly used in the US in 2020 (as of 1 October 2020) according to ahrefs.com.[2] From the original list of 100 terms, we removed 21 terms due to either being directly related to Google (such as “google” or “google maps”) or related to domains that are already used in the lists of preferred domains (such as “cnn” or “foxnews”), leaving 79 terms to be used in the generation process.

2) **Country-Specific Queries**: These terms are generated by combining country names with the random modifiers “culture”, “GDP”, “GDP per capita”, “history”, “neighbouring countries”, “people”, “population”, and “sports”.

3) **Unit Conversion and Arithmetic**: These include random unit conversions and simple arithmetic equations to diversify the set of search terms.

The generation process involves selecting and combining these terms using randomization to ensure that the resulting search queries are varied and cover a broad spectrum of topics. Algorithm 1 describes the core algorithm used in the script to generate the non-partisan search terms.

Daily throughout our study, this procedure generated a list of 117 non-partisan search terms, consisting of 79 common terms, 15 country-specific terms, 14 unit conversions, and 9 arithmetic equations. From these lists, the users visited a random subset of two search terms (during the early phase of the study) or one search term (during the later phase of the study) per day. Table E5 presents a random selection of 300 out of the 2,581 such terms generated and used in our study. All 2,581 terms generated and used in our study are made available in the replication package.

[1] The script is publicly archived here: `https://raw.githubusercontent.com/umatter/primemover_py/master/src/auxiliary/GenerateBenignTerms.py`.

[2] The archived copy of the original list can be found here: `https://web.archive.org/web/20201001170413/https://ahrefs.com/blog/top-google-searches/`.

Algorithm 1 Generate Non-Partisan Search Terms

```
procedure GENERATENONPARTISANSEARCHTERMS
    Load common_terms
    Load countries
    Initialize country_modifiers with predefined modifiers
    Initialize terms as an empty list
    for each term in common_terms do
        Add term to terms
    end for
    country_terms ← Random sample of 15 countries
    for each country in country_terms do
        mod ← Random choice from country_modifiers
        Add country + mod to terms
    end for
    for i = 1 to 14 do
        Add random unit conversion to terms
    end for
    for i = 1 to 9 do
        Add random equation to terms
    end for
    Save terms
end procedure
```

B4 Partisan search terms

We generate empirically reasonable lists of liberal and conservative search terms in four steps. First, following the procedure suggested in Gentzkow and Shapiro (2010), we compute the 'partisan loading' of phrases (bigrams) used in national US politics. To this end, we collect data on all phrases used by Members of Congress (MoC) in tweets and congressional speeches during the 116th US Congress (the period relevant for our study). Speech data is collected from the Congressional Record provided in digital form by the Library of Congress (see `https://www.congress.gov/congressional-record`); tweets are collected from the MoCs' Twitter feeds. We only use those bigrams considered "valid" in the sense of Gentzkow, Shapiro and Taddy (2019) (i.e., procedural bigrams, non-speech-related bigrams from the Congressional Record, etc. are removed). For all processing of text data described in this subsection, we use stemmed bigrams. We then compute the 'partisanship' of each (stemmed) phrase/bigram p as

$$\chi_p^2 = \frac{(f_{pr} f_{\sim pd} - f_{pd} f_{\sim pr})^2}{(f_{pr} + f_{pd})(f_{pr} + f_{\sim pr})(f_{pr} + f_{\sim pr})(f_{\sim pr} + f_{\sim pd})}, \tag{B2}$$

where f_{pr} and f_{pd} denote the total number of times bigram p is used by Republicans and Democrats, respectively, and $f_{\sim pr}$ ($f_{\sim pd}$) denote the total number of times a bigram that is not bigram p is used by Republicans (Democrats). A higher χ_p^2 value indicates that p is used predominantly by members of one of the two parties. For the next steps, we select the 500 most partisan phrases.

Second, we want to map these 500 bigrams to an ideology scale from -1 (clearly Democratic/liberal) to 1 (clearly Republican/conservative). To do so, we collect data on each MoC's ideological position from Voteview.[3] We denote their DW-Nominate score for MoC c by π_c. Next, we compute the relative frequency with which each MoC c uses a given phrase/bigram p: $\tilde{f}_{pc} = f_{pc}/\sum_{p=1}^{P} f_{pc}$. Again, closely following (55), we regress $\tilde{f}_{pc}$ on π_c for each bigram p, which gives us intercepts α_p and slope coefficients β_p. A positive β_p means that the more often a MoC uses p (relative to other terms), the more Republican/conservative she is. β_p thus indicates bigram p's location on the liberal-conservative scale. In the same vein, we interpret $SE(\beta_p)$ as an indication of whether the position of bigram p on the liberal-conservative scale is more or less precisely measured.[4] That is, we interpret a bigram p with a positive and large t-value of β_p as 'clearly conservative' and a bigram p with a large but negative t-value of β_p as clearly liberal. Finally, we rescale the t values of β_p to $[-1,1]$.

Third, we check whether the most partisan bigrams identified are used as search terms in Google. For each complete phrase that matches one of the stemmed bigrams p we verify whether, when, and in which region it was used as a search term on Google. We query the Google Trends (2020) platform for each of the completed (unstemmed) most partisan bigrams.[5] We then verify how often (in relative terms) each of the remaining unstemmed bigrams are used as search terms in Google. Google Trends provides relative search term frequencies, so-called "Interest" values, on a scale from 0–100, with 100 indicating the search term with the highest relative frequency of a maximum of five selected search terms in a given state (or other geographical unit) and period.[6] Therefore, the usage frequencies of search terms are always expressed relative to each other and are specific to a given state at a given time. As the Google Trends platform only allows comparisons of five search terms at a time, we query the search term frequencies in batches of four partisan bigrams and add to each batch a common reference search term. We use "carbon free" as a common reference search term.[7] By using a common

[3] Voteview (`https://voteview.com/data`) provides estimates of MoCs' ideological positions on a scale from -1 (very liberal) to 1 (very conservative). The ideological positions (so-called DW-Nominate scores) are inferred from roll call records, using the scaling method suggested by Poole and Rosenthal Poole and Rosenthal (1985).

[4] For example, a given phrase might be used often by rather moderate Republicans who are ideologically not so far from rather centrist Democrats. Now suppose these moderate Democrats and Republicans have relatively close but nevertheless clearly distinct positions on the liberal-conservative scale (according to their voting behavior, measured by DW-Nominate scores). It could be that a given phrase is used only slightly more often by moderate liberals than by moderate conservatives, and β_p is not statistically significantly different from 0. In such cases, $SE(\beta_p)$ helps us to take into consideration whether a phrase is clearly indicative of a MoC's position on the liberal-conservative scale.

[5] The 500 most partisan bigrams map to roughly 1,400 complete phrases. For example, the stemmed bigram "climat chang" maps to the complete phrases "climate change", "climate changing", "climate changes", "climate changed", "climatic change". Per stemmed bigram, we only keep the corresponding unstemmed partisan bigram that is most frequently used as a search term in Google.

[6] According to Google Trends (2020), Interest values from these time series are to be understood as representing "search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular."

[7] A good reference search term should satisfy the following criterion: The frequency with which it is used in the US should be relatively high and constant across regions/states and over the relevant period. We evaluated several (partisan) bigrams such as "access health," "carbon free," and "Latino vote." First, we verified in how many states and over how many years, each of these bigrams was used as a search term in the US. All candidate reference search terms were used in over 40 states and were searched rather constantly often over time in the period from October 2019 to October 2020, with "carbon free" being the term with the least variation in search frequency over time. Second, we ran comparisons of these candidate reference search terms with 300 randomly chosen partisan bigrams on Google Trends. We then counted how often Google Trends showed Interest values as "< 1". Having many "< 1" Interest values would make it harder to compare the relevance of partisan terms when used as search terms. Thereby, "carbon free" was the term with the least cases of "< 1" Interest values. Thus, this term is constantly relatively often used over time as well as across US states (and has roughly similar search volumes as many of the partisan bigrams). Thus, we chose "carbon free" as our reference search term.

reference search term, we ensure that all search term frequencies are relative to the same reference frequency (both across states and over time). We then only keep the bigrams that are used on at least 50 days and in at least 10 US states since 2016 as partisan search terms.

Fourth, we extend the compiled set of partisan search terms with search queries that, according to Google Trends (2020), are related to (substitutes for) our partisan search terms. To this end, we use information (provided by Google Trends, 2020) that relates other queries to a given search query (as long as the latter is used rather often).[8] We use the 'related queries' information to select for each partisan search term related queries that are searched for at least 90 percent as often as the corresponding original search term (using Google Trends' 'Top' metric). We can think of these related queries as alternative formulations/synonyms of the original partisan search terms. Following up on the example from above, the stemmed bigram "clean energi" is identified as one of the most partisan bigrams and is clearly identified as a typically liberal term (with a β_p t-value of -6.44, rescaled to $\pi_p = -0.477$). From all the unstemmed bigrams mapping to "clean energi", "clean energy" is the one most frequently used as a search term. Finally, based on queries related to "clean energy", we can map "clean energi" to the synonymous search terms "solar energy", "clean renewable energy", "renewable energy", and "clean energy". Search terms suggested by Google Trends' 'related queries' that were unrelated to the original term or almost identical to the original term were manually removed.

Out of the roughly 1,200 partisan search terms identified in this way, we select the 400 most clearly partisan search terms, label the conservative (liberal) ones as Republican (Democrat) search terms, and use them as the basis for the users' search vocabulary.[9] We randomly assign to each Republican (Democrat) user a set of 10 Republican (Democrat) search terms. In addition, we assign each Republican (Democrat) user a set of 10 highly partisan (and election-related) hashtags to be used as additional search terms (collected from Best Hashtags, 2020*a*,*b*). Tables E6 and E7 list the slant-based and the hashtag-based terms assigned to Democratic and Republican users, respectively. The higher number of terms assigned to Democratic users follows from the fact the majority of highly partisan search terms were identified as having a liberal slant.

Finally, partisan users also directly use the domain names of their ten preferred partisan domains as search terms (instead of typing the entire domain into the browser bar).

B5 Validation of partisan search terms

To validate our final selection of partisan search terms, we use the time-averaged relative frequency $\bar{f}_{ps}$ with which a given partisan search term p is used for Google queries from computers located in state s. We compute this frequency based on Google Trends (2020) queries. Remember that the

[8] Specifically, "[u]sers searching for [this] term also searched for these [related] queries" (Google Trends, 2020).

[9] Specifically, we select the top 400 cases with an absolute value of π_s greater than 0.5, and label search terms with negative β_p values as Democratic and those with positive values as Republican.

Google Trends platform only provides relative search term frequencies, so-called Interest values, for up to five search terms at a time. Hence, in response to any query with five search terms, the Google Trends platform provides an Interest value h_{ist} for each search term i. As we use four partisan search terms and our common reference search term r ("carbon free") in each of our Google Trends queries, we can compute the relative frequency f_{pst} of partisan search term i in state s and time period t as $f_{pst} = h_{pst}/h_{rst}$. Subsequently, we compute the time-averaged relative frequency of partisan search term p in state s as $\bar{f}_{ps} = \frac{1}{T}\sum_{t=1}^{T} f_{pst}$.

We then compute measures for the popularity of Republican and Democratic search terms in the different states. First, we compute the Republican search volume in state s as

$$RepSearches_s = \frac{\sum_{p\in R}\bar{f}_{ps}}{|R|\times 100}, \tag{B3}$$

where R is the set of search terms labeled as Republican. Second, we analogously compute $DemSearches_s$ based on the set of search terms labeled as Democratic. Finally, we compute the net Republican search volume in state s as $RepSearches_s - DemSearches_s$.

In the last step, we compare the net Republican search volume with the net share of Republican votes in the 2020 US election (as a proxy for the locally dominant ideology) across states. Figure D5 shows a positive correlation (with the raw correlation coefficient being 0.49), suggesting that a higher share of Republican voters in a given state tends to be reflected in the search behavior of this state's population. In the aggregate, Google users in more conservative states tend to use conservatively rather than liberally slanted search terms. These validation results at the aggregate level are consistent with recent experimental evidence on partisan information seeking at the individual level (Peterson and Iyengar, 2021).

B6 Election-related search terms

Before, during, and after the elections, Google Trends (2020, 2021) curated special feature pages summarizing search trends related to the US 2020 election. In particular, Google Trends kept track of how frequently the presidential candidates' names were searched, and it kept updating a ranked list of the 25 most trending search terms on the election (for the respective past seven days). Figure D6 shows a screenshot of one of these feature pages focusing specifically on the elections and voting on 1 November 2020. Later similar feature pages were published on the presidential transition, the inauguration, and Biden's new cabinet.

Throughout our study, we collected the most trending election-related searches from these feature pages on Google Trends (2020, 2021) every week and selected a small subset of the most highly ranked search terms from these lists, thereby preferring rather specific over overly generic terms.[10] On each day, we then chose two of these election-related search terms and assigned them to all users. Table E8 lists all election-related search terms used at some point in our study.

[10] For example, if both "election results" and "election results 2020" were very highly ranked in the same week, we would only select the second, more specific term, so that we could select an additional lower-ranked term and get a more diverse set of search terms. Similarly, if both "voting" and "did my vote count?" were very highly ranked, we would only select the second, more specific term.

C Analysis of Search Results

C1 Web search personalization: Robustness tests

Panels C and D of Figure 2 in the main text present our main results on behavior- and location-based search result personalization. They are based on the regression outputs in Table E10 (column 1) and Table E11 (column 1), respectively. Here we discuss various robustness tests.

Figure D10 is similar to panel C of Figure 2 but looks at the effect of previous visits to some specific preferred domains on the probability that these domains occur on the first search results page in response to the election-related queries. We thereby focus on the 20 preferred partisan and popular domains that occurred most often in the users' election-related search results. More specifically, we run a separate regression for each of these 20 domains, in which we regress an indicator variable for the occurrence of this domain in a user's election-related search results on the number of the user's previous visits to this domain, thereby accounting for date of search fixed effects and two-way clustering the standard errors by user and date of search. We see considerable heterogeneity in the estimated effect sizes. The five domains with the largest effect sizes include three national news domains, an information platform, and an official state government domain.

Table E10 shows more robustness tests for the results reported in panel C of Figure 2. Thereby, we again impose a linear relationship between the number of previous visits of preferred domains and the number of preferred domains on the first search results page in all specifications from column 2 onward. Moreover, from column 4 onward, we change from a daily panel with user $\times$ day as units of observations to a panel with user $\times$ election-related searches as units of observations. Columns 2–5 show results for these two panels with varying fixed-effects specifications. Column 6 is identical to column 5, but we restrict the sample to search results that contain at least one of the domains preferred by the user. Finally, we look at alternative dependent variables. In column 7, we use a simple indicator variable that is equal to 1 if the search results contain at least one domain from the user's set of preferred domains and 0 otherwise. In column 8, the dependent variable is the rank of the (highest-ranked) preferred domain (and we again restrict the sample to search results that contain at least one of the user's preferred domains). The estimate shows that, conditional on the search results containing at least one preferred domain, the more often a user has previously visited preferred domains, the higher up in the search results those domains tend to occur.

Tables E11 and E12 present robustness tests for the results on location-based personalization shown in panel D of Figure 2. In columns 2–5 of Tables E11, we replace the explanatory variables and use the number of users in the city that see the corresponding domain. These columns present the results for the full sample, only Democratic users, only Republican users, and only non-partisan users. Columns 5–8 present the same analyses at the level of states rather than cities. In addition, Table E12 shows results from regressing an indicator equal to 1 if a specific domain occurs on a user's first search results page on an indicator equal to 1 if the majority of other users located in the same city (columns 1–3) or state (columns 4–6) see this domain in their search results. Columns 1 and 4 show a baseline specification. Columns 2 and 5 add domain fixed effects and clustering of standard errors at the level of domains. Columns 3 and 6 re-estimate the baseline specification with a logistic regression model. Estimation of the logistic regression with fixed effects is implemented based on the approach suggested by Stammann (2018). All these robustness tests corroborate our finding of location-based personalization.

C2 Search result ideology: Details on ideology indices and robustness tests

We use the following domain ideology indices (presented in alphabetical order) to compute the Search Results Ideology Score (SRIS) defined in equation (3) in the main text.

- Alignment score: Bakshy, Messing and Adamic (2015) propose a partisan alignment score that indexes domains on a continuous scale from -1 (liberal) to 1 (conservative) based on the relative frequency with which webpages of these domains are shared on Facebook by self-identified liberal or conservative Facebook users.

- Partisanship score: Budak, Goel and Rao (2016) propose a partisanship score based on the representation of the Republican Party and the Democratic Party in political news articles. The score indexes (online) news outlets on a scale from -1 (left-leaning) to 1 (right-leaning).

- MTurk bias score: Robertson et al. (2018) use human raters on MTurk to code a subset of domains used in their main index (see below) on a five-point Likert scale from -1 (liberal) to 1 (conservative).

- Pew Research Center score: Mitchell et al. (2014) from the Pew Research Center use a survey with several policy-related questions to map survey participants on a five-point scale from consistently liberal to consistently conservative and study which news outlets the respondents trust most. Based on these survey data, Robertson et al. (2018) create an index on a liberal-conservative scale from -1 to 1, reflecting which online news outlets tend to be trusted by liberals/conservatives. We use the index provided by Robertson et al. (2018).

- Partisan audience bias score: Robertson et al. (2018) propose a partisan audience bias score based on the shared web domains by registered Democratic and Republican voters on Twitter. The score scales from -1 (if the domain is exclusively shared by registered Democrats) to 1 (if the domain is exclusively shared by registered Republicans).

Figure 3 in the main text present our main results on the effects of the users' partisanship and the cities' partisan leanings on the SRIS. This figure is based on the regression outputs in columns 1–4 of Table E13. We now discuss various robustness tests.

Figure D11 replicates Figure 3 but only based on the observations during the first half of the study period. We find that the results look similar to our main results. However, we lose some power, leading to larger confidence intervals in some instances.

Columns 5–9 of Table E13 provide robustness tests based on a more granular dataset in which we count the occurrences of preferred domains at the level of election-related searches instead of per day. This allows us to use more stringent fixed effects specifications. The unit of observation is thus user $\times$ first search results page. Based on the same unit of observation, columns 9–13 then document our main results' sensitivity to replacing the usual SRIS (based on the average of several ideology indices) with the values of individual domain ideology indices. That is, we compute the SRIS for each ideology index separately and regress the resulting score on our explanatory variables. Although, not surprisingly, the results vary from index to index, the overall picture is qualitatively consistent with specifications based on the aggregate index.

Table E14 uses more fine-grained explanatory variables, specifically the average ideology score of the users' previously visited preferred domains and the city-level share of Republican votes. The results confirm that the prevalent partisanship of the users' location is more important than the partisanship of their browsing and search history.

D Additional Figures

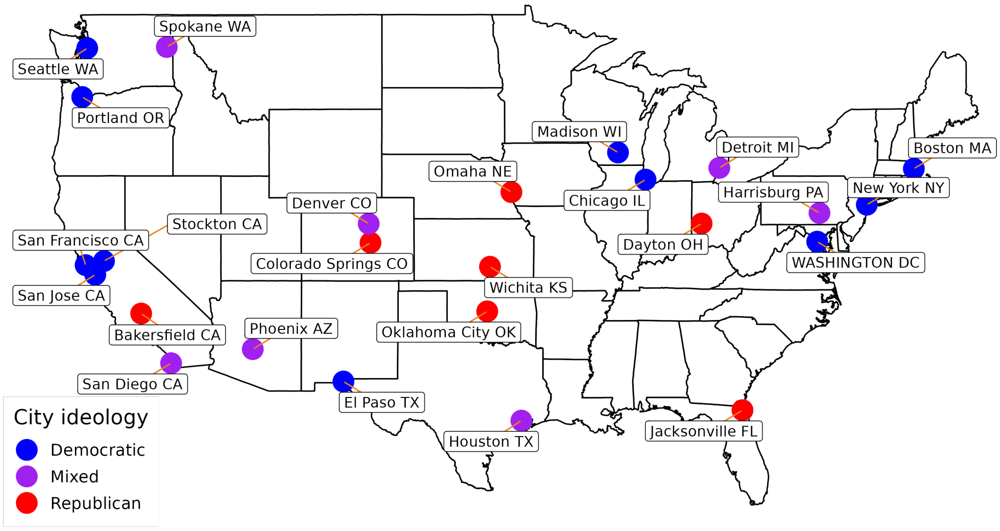
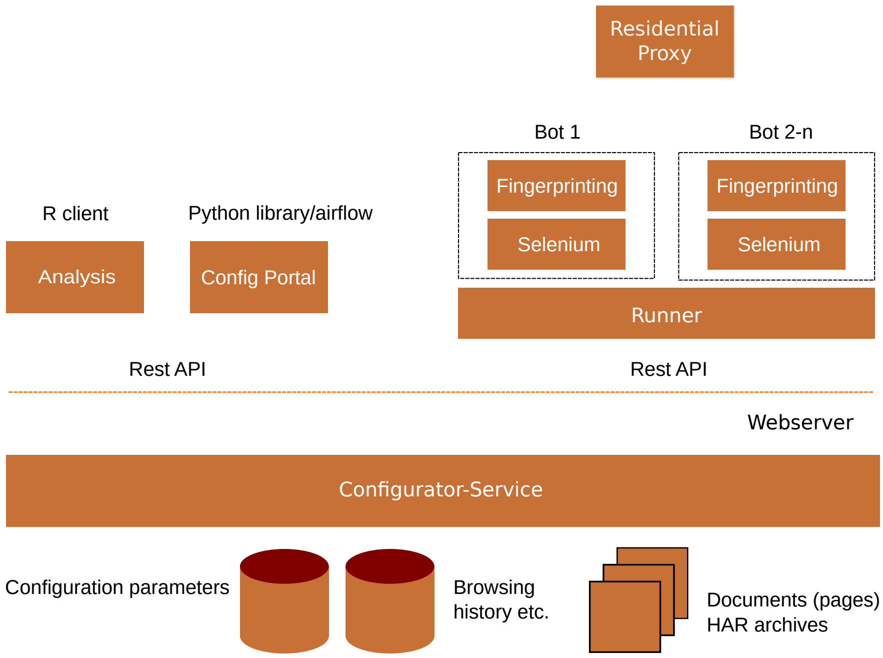


FIGURE D1. SPATIAL DISTRIBUTION OF USERS AND LOCATION/CITY IDEOLOGY

Notes: The map shows the spatial distribution of the 150 synthetic users. The 25 cities include Democratic strongholds (blue), Republican strongholds (red), and "purple" cities, where neither party dominates. The city categorization is based on the Republican vote share in the 2016 US Presidential Elections (see Table E1). In each city, there are two Democratic, two Republican, and two non-partisan users, each with randomized differences in appearance and behavior.

FIGURE D2. ILLUSTRATION OF THE BASIC SOFTWARE ARCHITECTURE

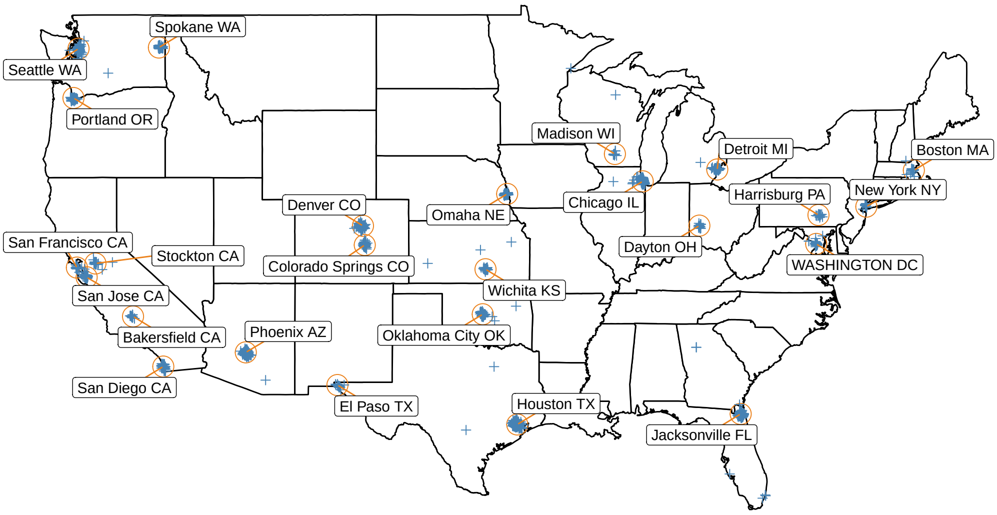


Notes: The software consists of three core components: The "runner" instance, which handles and runs the synthetic users' browsing and search sessions, the configurator-service (API), which stores and provides all user configurations, jobs, and recorded web traffic (in HAR files), as well as the configuration portal (a client library to set up and configure the user population).

FIGURE D3. FINGERPRINTING VERIFICATION EXAMPLE

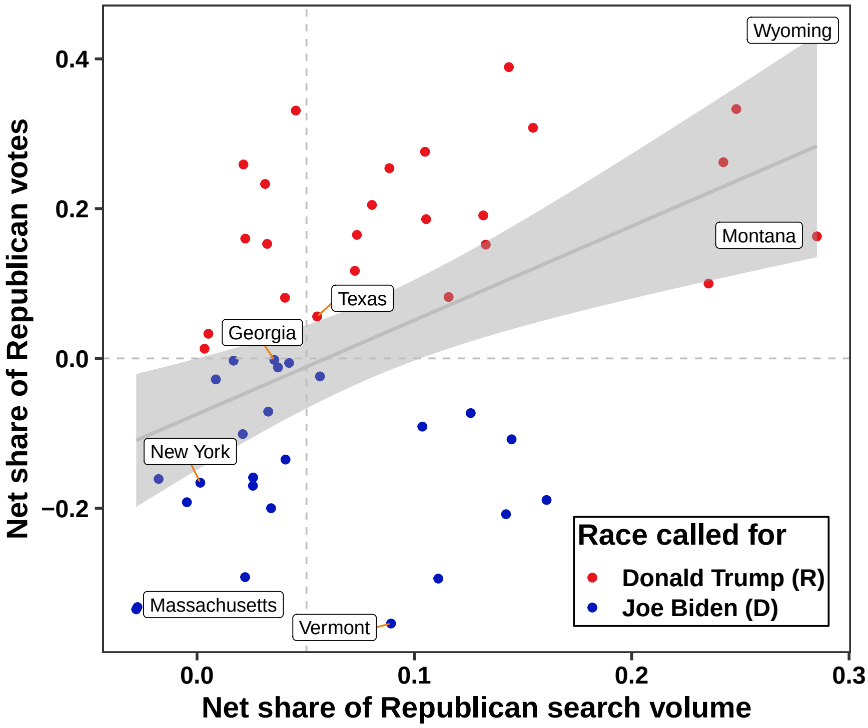
Notes: Sample screenshot showing the first few rows of a synthetic user's JavaScript browser profile on `www.browserleaks.com`. The example illustrates that the user's JavaScript settings are correctly recognizable and functional as well as that the synthetic user's (virtual) screen is properly recognized (screen properties recognition and canvas fingerprinting are some of many techniques used to identify and track users based on their device characteristics).

FIGURE D4. GEOLOCATION VERIFICATION



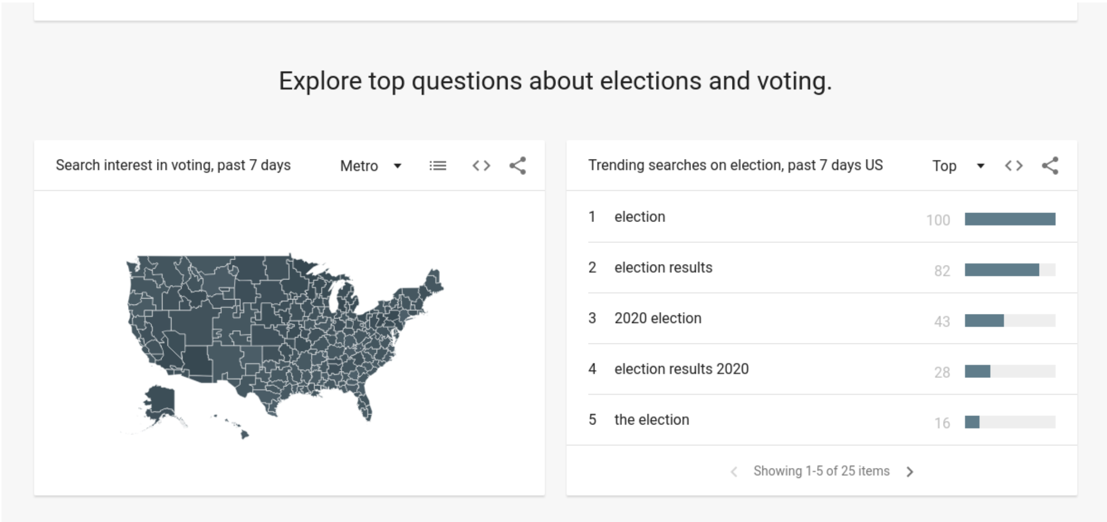
Notes: Mapping of unique coordinates extracted during synthetic user geolocation verifications from all browsing sessions (based on the third-party real-time geolocation service IPStack). Blue crosses indicate the verified coordinates of users when browsing and searching (all browsing sessions of all users included), and orange circles highlight the official city coordinates to which synthetic users were assigned during the study.

FIGURE D5. VALIDATION OF PARTISAN SEARCH TERMS: REPUBLICAN GOOGLE SEARCH VOLUME AND REPUBLICAN VOTE SHARES PER STATE



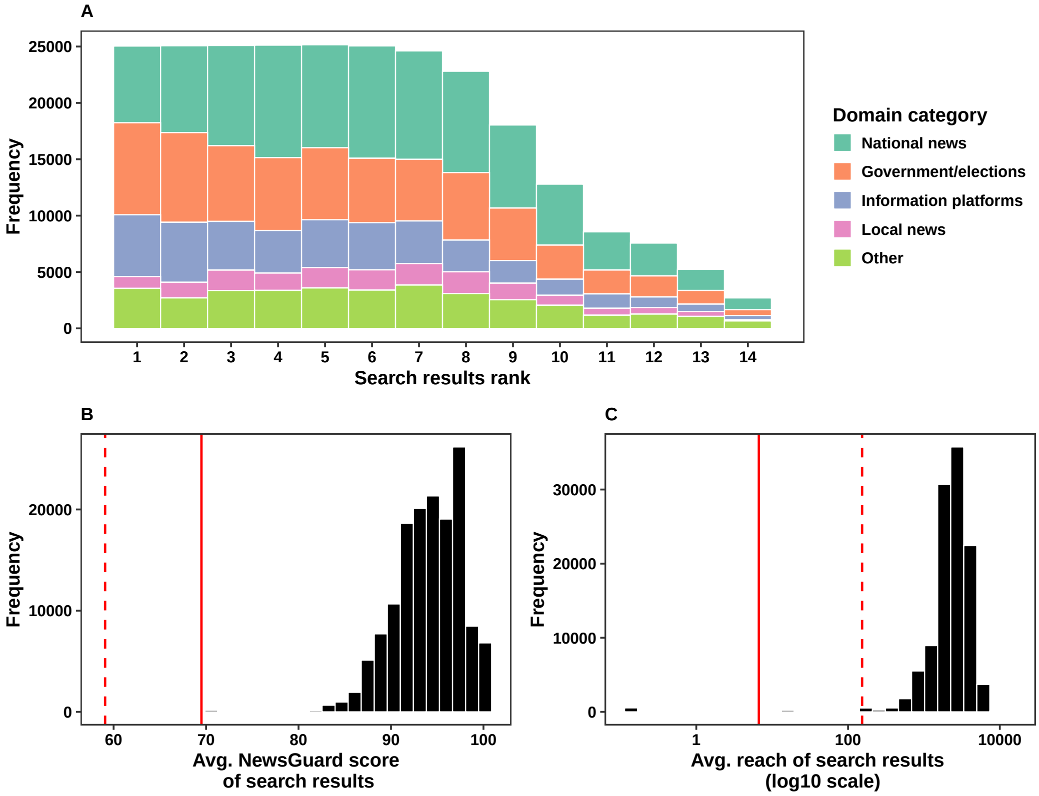
Notes: The scatter plot displays the raw correlation between the ideological position of a state based on its population's use of partisan search terms (the net Republican search volume as defined in Section B4 of the Online Appendix) and the corresponding state's net share of Republican votes (i.e., the share of Republican votes minus the share of Democratic votes according to the Federal Election Commission, 2022) in the 2020 US presidential elections. Dots of states where the US presidential race was called for Joe Biden are blue, those where the race was called for Donald Trump are red. The gray line indicates the intercept and slope coefficient from regressing the net share of Republican voters on the net Republican search volume. The gray area indicates the corresponding confidence band at a 95 percent confidence level. The dashed vertical line indicates the median net Republican search volume, and the horizontal dashed line indicates a net Republican vote share of 0. The election results data is from the Federal Election Commission (FEC). Data on search volume are collected from Google Trends (2020) (as described in Section B4).

FIGURE D6. GOOGLE TRENDS ELECTION FEATURE PAGE



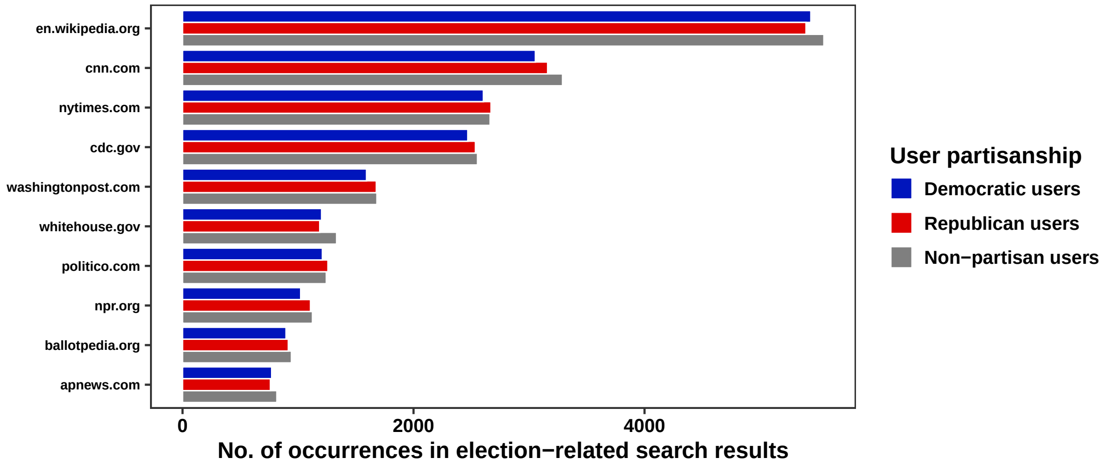
Notes: Exemplary screenshot showing a section with trending election-related search terms on `https://trends.google.com` on November 1, 2020. The panel on the right is one of the data sources used to collect the election-related search terms assigned to synthetic users.

FIGURE D7. CATEGORIZATION AND DESCRIPTION OF THE CONTENT OF ELECTION-RELATED SEARCH RESULTS



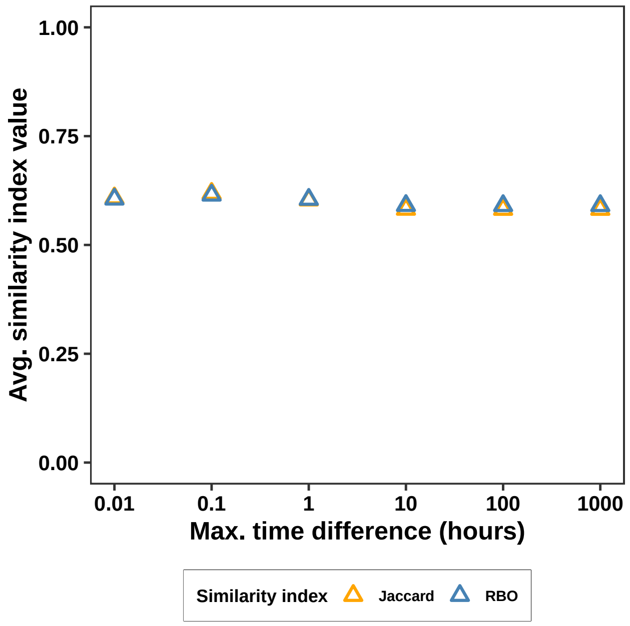
Notes: The histogram in panel A shows the rank distributions of search results per category of domains according to the corresponding dataset in Matter and Hodler (2024). (*Other* includes foreign news domains, business domains, and non-categorizable domains). Panel B shows the distribution of the average NewsGuard (2023) trust score of the domains on the search results pages (0–100 scale) vs. the median (red) and mean (dashed red) average NewsGuard trust score for all domains. The average NewsGuard trust score per domain is computed based on all NewsGuard assessments for this domain during the time period of our study. Panel C shows the histogram of the average search result reach (visitors per 1M web users from Amazon Web Information Services) vs. the median (red) and mean (dashed red) reach for 1,600+ US information platforms and news domains.

FIGURE D8. DOMAINS OFTEN OCCURING IN SEARCH RESULTS



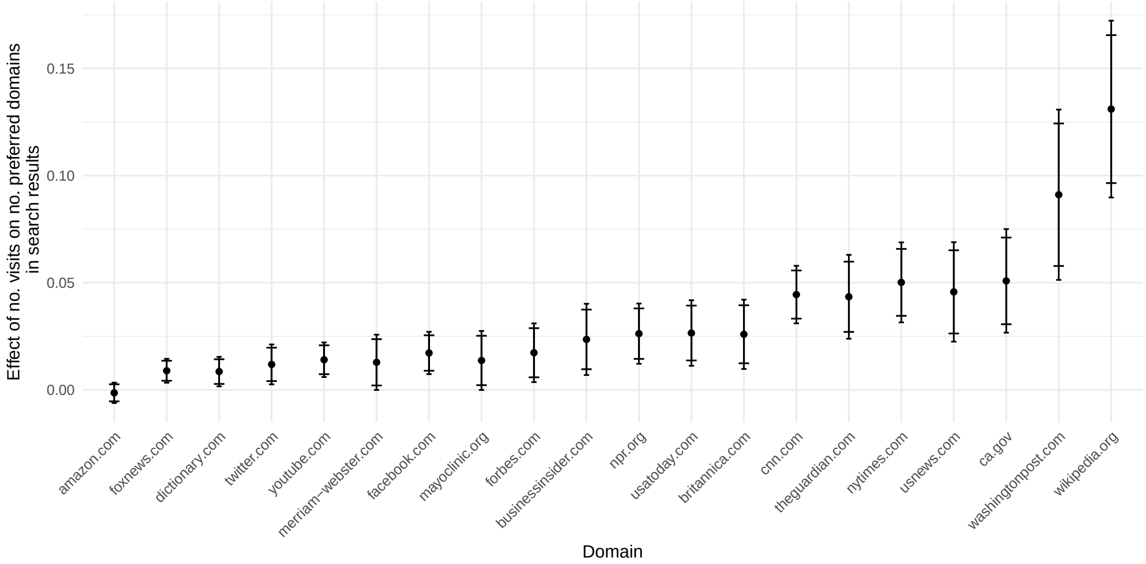
Notes: The barplot shows the number of occurrences of the ten most often occurring domains in election-related search results for Democratic users (blue bars), Republican users (red bars), and non-partisan users (grey bars).

FIGURE D9. SEARCH RESULTS SIMILARITY AND THE TIME DIFFERENCE BETWEEN SEARCHES



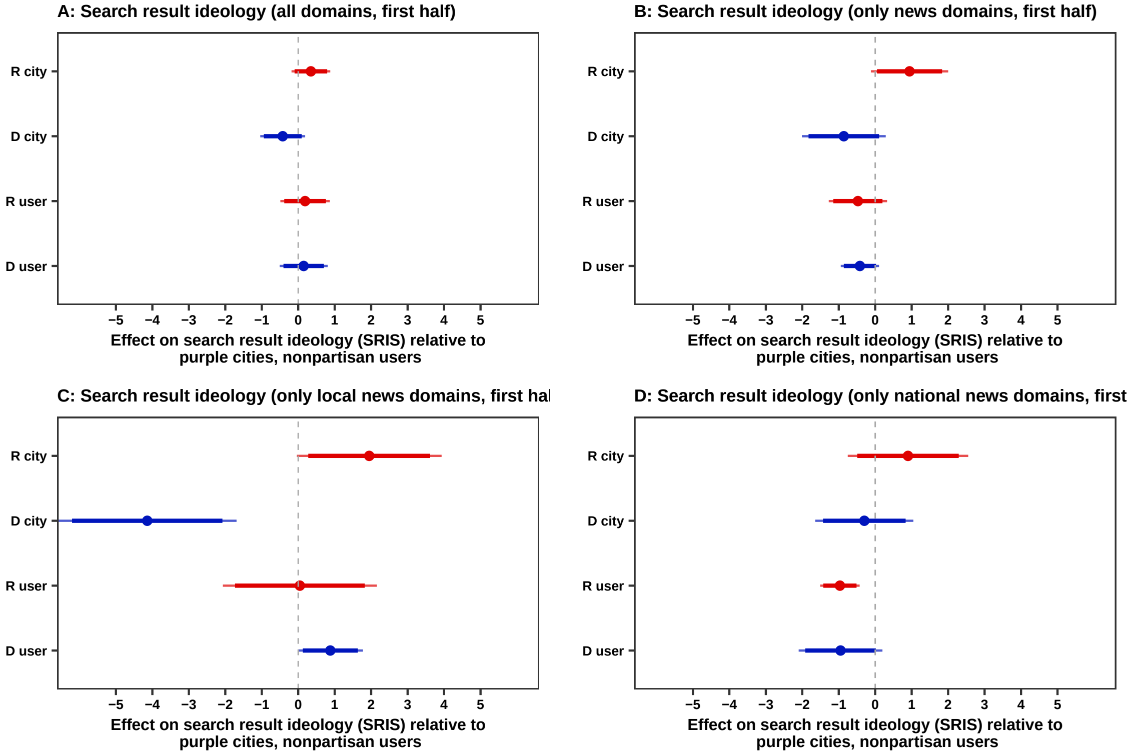
Notes: The plots display the average search results similarity (Jaccard and extrapolated RBO indices) for different search results-pair sub-samples. Sub-samples are generated based on a varying maximum duration threshold (in hours) between the queries of user i and user j using the same election-related search term. The thresholds to create subsamples are indicated on the horizontal axis; the corresponding average index values are indicated on the vertical axis. Reading example: the average extrapolated RBO similarity between search results resulting from Google queries using the same search term within 0.01 hours or less is 0.61.

FIGURE D10. DOMAIN-SPECIFIC BEHAVIOR-BASED PERSONALIZATION



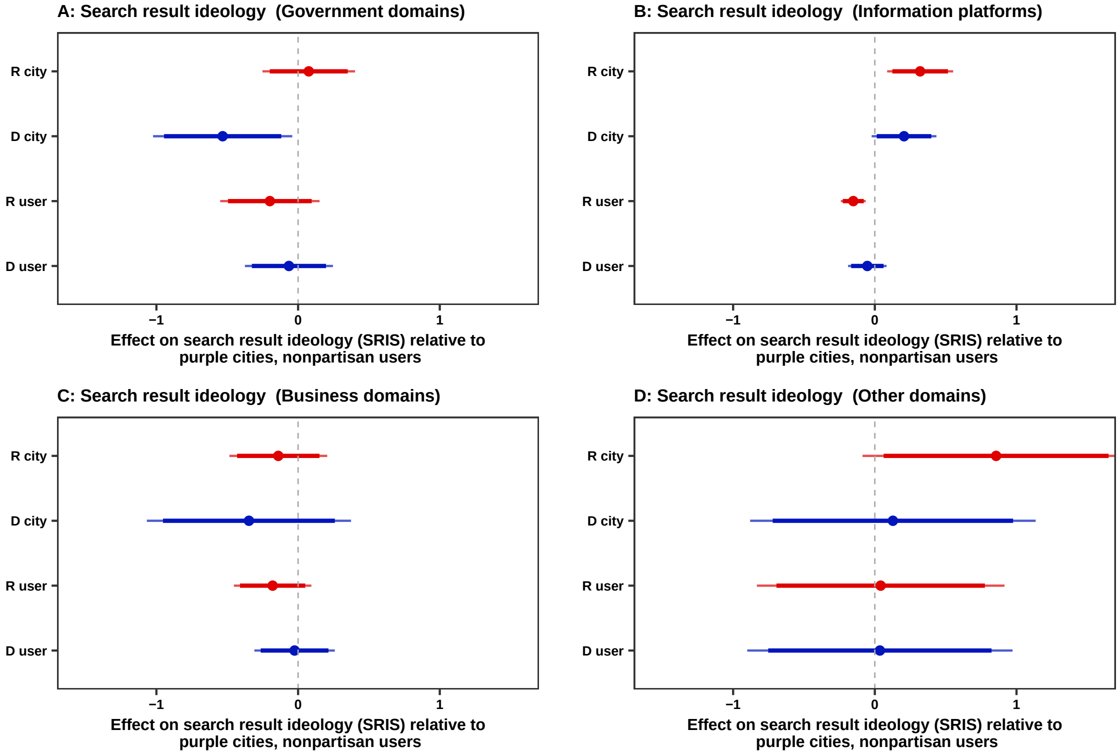
Notes: This figure complements panel C of Figure 2 by showing domain-specific personalization effects for the 20 preferred partisan and popular domains that occurred most often in the users' election-related search results. It shows the results from separate regressions for each of these 20 domains, in which we regress an indicator variable for the occurrence of this domain in a user's election-related search results on the number of the user's previous visits to this domain, accounting for date of search fixed effects. The 90% (wide vertical lines) and 95% (narrow vertical lines) confidence intervals are based on standard errors two-way clustered by user and date of search.

FIGURE D11. EFFECTS OF PARTISANSHIP AND LOCATION ON THE IDEOLOGY OF ELECTION-RELATED SEARCH RESULTS IN THE FIRST HALF OF THE STUDY PERIOD



Notes: Replication of Figure 3 based on all observations falling into the first half of the study period (i.e., prior to 16 December 2020). See the notes to Figure 3 for further information.

FIGURE D12. EFFECTS OF PARTISANSHIP AND LOCATION ON THE IDEOLOGY OF ELECTION-RELATED SEARCH RESULTS OTHER THAN NEWS



Notes: This figure complements Figure 3 by showing the effects of the users' partisanship and their cities partisan leaning on the SRIS based on domains in four non-news domain categories: government domains, information platforms, business domains, and other domains. See the notes to Figure 3 for further information.

E Additional Tables

TABLE E1—SYNTHETIC USER LOCATIONS AND CITY IDEOLOGY

City (user location)	GOP vote share (in %)	City ideology
CA-SAN FRANCISCO	13.40	D
WI-MADISON	20.00	D
CA-SAN JOSE	20.90	D
DC-WASHINGTON	23.10	D
TX-EL PASO	26.50	D
WA-SEATTLE	28.50	D
IL-CHICAGO	29.60	D
OR-PORTLAND	30.40	D
CA-STOCKTON	31.40	D
MA-BOSTON	32.20	D
NY-NEW YORK	33.30	D
CA-SAN DIEGO	37.90	purple
CO-DENVER	39.00	purple
MI-DETROIT	41.70	purple
TX-HOUSTON	44.20	purple
PA-HARRISBURG	49.80	purple
AZ-PHOENIX	50.70	purple
WA-SPOKANE	51.10	purple
OH-DAYTON	53.60	R
NE-OMAHA	53.90	R
FL-JACKSONVILLE	54.20	R
CA-BAKERSFIELD	55.00	R
KS-WICHITA	57.20	R
OK-OKLAHOMA CITY	58.90	R
CO-COLORADO SPRINGS	59.60	R

Notes: List of all US cities where synthetic users were located via residential proxies. The middle column shows the city-level Republican vote shares (in percent) in the 2016 elections, and the right column the corresponding categorization into Democratic, Republican, and ‘purple‘ cities. *Data:* City-level GOP vote shares from Dottle (2019).

Table E2—: Top 100 most popular domains visited by users

Domain	Reach	No. Occ. D	No. Occ. R	No. Occ. NP
google.com	608,726	916	841	883
youtube.com	319,700	283	291	292
facebook.com	66,967	898	870	966
wikipedia.org	52,461	8,378	8,377	8,520
yahoo.com	51,573	7	10	8
amazon.com	44,430	264	273	252
reddit.com	30,080	0	1	0
live.com	29,760	0	0	0
netflix.com	28,030	0	0	0
microsoft.com	26,298	0	0	0
office.com	16,225	0	0	0
instagram.com	15,136	11	4	7
apple.com	12,592	135	127	120
ebay.com	10,893	0	0	0
msn.com	9,168	1	0	0
twitter.com	8,229	814	775	811
cnn.com	8,105	5,484	5,625	5,756
linkedin.com	6,980	0	1	0
nytimes.com	6,773	5,406	5,608	5,511
imdb.com	6,510	1	2	2
etsy.com	5,500	0	0	0
espn.com	5,420	61	58	62
spotify.com	5,200	0	0	0
walmart.com	4,518	0	0	0
indeed.com	4,465	0	0	0
paypal.com	4,110	0	0	0
theguardian.com	4,024	913	892	906
nih.gov	3,940	12	23	17
chase.com	3,610	0	0	0
foxnews.com	3,400	523	544	543
bestbuy.com	3,250	0	0	0
zillow.com	3,250	0	0	0
cnet.com	3,227	132	156	138
quizlet.com	3,160	0	0	0
washingtonpost.com	3,108	3,090	3,181	3,190
pinterest.com	3,000	0	0	0
craigslist.org	2,605	0	0	0
wikihow.com	2,340	0	0	0
target.com	2,265	0	0	0
healthline.com	2,240	40	39	38
hulu.com	2,225	0	0	0
homedepot.com	2,160	0	2	0
investopedia.com	2,160	1	2	0
wellsfa.go.com	2,090	0	0	0
businessinsider.com	2,078	464	476	453
usps.com	1,980	0	0	0
forbes.com	1,925	267	273	271
weather.com	1,792	0	1	0
steamcommunity.com	1,780	0	0	0
ca.gov	1,594	963	973	973
gamepedia.com	1,570	0	0	0
usatoday.com	1,497	811	836	866
aol.com	1,444	3	2	5
quora.com	1,410	1	0	0
xfinity.com	1,390	0	0	0
webmd.com	1,270	81	78	82
yelp.com	1,270	0	0	1
npr.org	1,230	1,867	1,944	1,998
wayfair.com	1,230	0	0	0
steampowered.com	1,220	0	0	0
genius.com	1,200	0	0	0
tripadvisor.com	1,160	1	1	0

(Continued on next page)

Domain	Reach	No. Occ. D	No. Occ. R	No. Occ. NP
glassdoor.com	1,150	0	0	0
techradar.com	1,125	0	0	0
bankofamerica.com	1,067	0	0	0
merriam-webster.com	1,000	232	232	213
cbssports.com	978	31	32	30
britannica.com	976	560	539	540
lowes.com	972	0	0	0
mayoclinic.org	909	212	213	224
wiktionary.org	888	0	0	0
macys.com	864	0	0	0
ign.com	848	0	0	0
accuweather.com	839	0	0	0
usnews.com	820	421	452	475
dictionary.com	788	255	260	279
huffpost.com	782	145	121	140
urbandictionary.com	752	5	3	2
irs.gov	735	10	19	8
rottentomatoes.com	669	0	2	0
medicalnewstoday.com	627	0	0	0
allrecipes.com	611	0	0	0
bleacherreport.com	535	5	2	2
com.go.com	530	0	0	0
expedia.com	524	0	0	0
groupon.com	461	0	0	0
foodnetwork.com	375	0	0	0
bbb.org	354	0	0	0
mapquest.com	324	0	0	0
apartments.com	301	0	0	0
fb.com	135	9	9	8
fandom.com	NA	35	33	26
retailmenot.com	NA	0	0	0
roblox.com	NA	0	0	0
wowhead.com	NA	0	0	0
yellowpages.com	NA	0	0	0

Notes: The table lists all most popular (canonical) domains randomly visited by users during our study. The first column shows the domain names. "Reach" refers to the reach (visitors per 1M web users) provided by AWIS (2020). Domains for which this data is missing in our dataset are indicated with NA. The remaining three columns indicate how often a domain occurs in users' election-related first search results pages during our study, separately by type of user (Democratic, Republican, non-partisan). The table is ordered by "Reach". Section B1 describes the selection of these domains.

Table E3—: TOP DEMOCRATIC DOMAINS ASSIGNED TO DEMOCRATIC USERS

Domain	Partisanship	Reach	No. Occ. D	No. Occ. R	No. Occ. NP
yahoo.com	0.17	51,573	7	10	8
msn.com	0.52	9,168	1	0	0
cnn.com	1.05	8,105	5,484	5,625	5,756
nytimes.com	0.79	6,773	5,406	5,608	5,511
medium.com	0.25	5,630	0	0	0
washingtonpost.com	0.68	3,108	3,090	3,181	3,190
usatoday.com	0.12	1,497	811	836	866
npr.org	0.31	1,230	1,867	1,944	1,998
fivethirtyeight.com	0.10	1,070	182	195	184
nbcnews.com	0.57	1,038	2,178	2,180	2,164
politico.com	0.47	922	2,450	2,501	2,473
huffpost.com	0.84	782	145	121	140
cbsnews.com	0.13	702	1,081	1,140	1,113
apnews.com	0.15	666	1,494	1,497	1,502
thehill.com	0.73	603	570	547	600
theatlantic.com	0.54	514	53	41	46
thedailybeast.com	0.15	464	318	353	333
latimes.com	0.10	438	263	260	236
vox.com	1.48	422	351	368	368
gofundme.com	0.14	401	0	0	0
newsweek.com	0.23	395	17	22	20
mashable.com	0.36	379	0	0	0
pbs.org	0.24	372	291	306	307
zazzle.com	0.18	339	0	0	0
slate.com	0.62	324	3	1	3
msnbc.com	0.37	314	2	0	1
iheart.com	0.12	255	0	0	0
esquire.com	0.21	252	0	0	0
nymag.com	0.13	226	136	140	146
vulture.com	0.20	213	0	0	0
tumblr.com	0.25	212	0	0	0
salon.com	0.15	180	10	8	11
snopes.com	0.16	176	54	40	46
actblue.com	1.41	149	0	0	0
dailykos.com	2.61	116	0	0	0
rawstory.com	0.80	116	0	0	0
boingboing.net	0.80	98	0	0	0
dailydot.com	0.67	93	0	0	0
theintercept.com	0.30	81	0	0	0
motherjones.com	0.36	74	40	34	23
jezebel.com	0.25	69	0	0	0
actionnetwork.org	0.16	62	0	0	0
talkingpointsmemo.com	0.83	59	0	0	0
theroot.com	0.13	48	0	0	0
amnesty.org	0.44	47	0	0	0
thenation.com	0.18	45	20	30	18
commondreams.org	0.42	37	0	0	0
jacobinmag.com	0.12	35	0	0	0
aclu.org	0.15	34	0	0	0
alternet.org	0.35	28	0	0	0
moveon.org	0.38	24	0	0	0
abovethelaw.com	0.17	22	0	0	0
weaselzippers.us	0.15	22	0	0	0
crooksandliars.com	0.74	18	0	0	0
mcsweeneys.net	0.11	16	0	0	0
politicususa.com	1.89	16	0	0	0
themarysue.com	0.25	16	0	0	0
democraticunderground.com	0.31	12	0	0	0
thinkprogress.org	0.67	7	0	0	0
amnesty.org.uk	0.12	5	0	0	0
nationalmemo.com	0.51	5	0	0	0
splinternews.com	1.78	4	0	0	0

(Continued on next page)

Domain	Partisanship	Reach	No. Occ. D	No. Occ. R	No. Occ. NP
bipartisanreport.com	0.14	3	0	0	0
opednews.com	0.51	3	0	0	0
theblacksphere.net	0.15	3	0	0	0
truthdig.com	0.27	3	0	0	0
itsgoingdown.org	0.15	2	0	0	0
trofire.com	0.27	2	0	0	0
unicornriot.ninja	0.11	1	0	0	0
credoaction.com	0.22	0	0	0	0
vote.org	0.13	NA	298	278	295
buzzfeednews.com	0.11	NA	26	24	32
amplifr.com	0.40	NA	0	0	0
bikudo.com	0.10	NA	0	0	0
corrupt.af	0.14	NA	0	0	0
doinmytoons.blogspot.com	0.10	NA	0	0	0
flake.news	0.55	NA	0	0	0
hannahsix.blogspot.com	0.33	NA	0	0	0
hillreporter.com	0.51	NA	0	0	0
iwillvote.com	0.10	NA	0	0	0
jakehasablog.blogspot.com	0.22	NA	0	0	0
mobilize.us	0.20	NA	0	0	0
odaction.com	0.45	NA	0	0	0
pinterest.ch	0.24	NA	0	0	0
qoo.ly	0.32	NA	0	0	0
usaunify.org	0.16	NA	0	0	0

Notes: The table lists all (canonical) domains randomly assigned to Democratic users out of the top 100 most Democratic domains identified based on the procedure described in Section B2 of the Online Appendix. The first column shows the domain names. The second column displays the corresponding value of the "Partisanship" measure, which indicates how partisan a domain identified as Democratic is. This measure is based on the χ^2 test outlined in Section B1 of the Online Appendix. The corresponding χ_i^2 is a test statistic for the null hypothesis that the propensity to mention domain i in a tweet is equal for Democrats and Republicans (based on a sample of millions of tweets from Republican and Democratic Twitter users during the 2018 midterm elections). The higher the χ_i^2 value is, the more significantly a domain is primarily mentioned by supporters of one of the two parties. For all domains identified as highly partisan in this way, we look at whether they were more often used by Democrats or Republicans. In the table here, we only list domains that are identified as highly partisan and were more often used by Democrats, and we multiply the corresponding χ_i^2 value by $1,000$ for ease of readability. "Reach" refers to the reach (visitors per 1M web users) provided by AWIS (2020). Domains for which this data is missing in our dataset are indicated with NA. The remaining three columns indicate how often a domain occurs in users' election-related first search results pages during our study, separately by type of user (Democratic, Republican, non-partisan). The table is ordered by "Reach".

Table E4—: TOP REPUBLICAN DOMAINS ASSIGNED TO REPUBLICAN USERS

Domain	Partisanship	Reach	No. Occ. D	No. Occ. R	No. Occ. NP
twitter.com	0.10	8,229	814	775	811
foxnews.com	0.11	3,400	523	544	543
cnbc.com	0.14	2,120	404	419	434
dailymail.co.uk	0.13	2,080	5	3	3
rt.com	0.17	1,540	0	0	0
bloomberg.com	0.09	1,392	301	290	284
marketwatch.com	0.10	1,061	48	58	54
elpais.com	0.14	996	0	1	3
spiegel.de	0.09	698	1	1	0
breitbart.com	0.17	602	0	0	0
sputniknews.com	0.18	590	0	0	0
ft.com	0.19	469	30	32	23
zerohedge.com	0.13	367	0	0	0
welt.de	0.08	331	0	0	0
aa.com.tr	0.12	324	0	2	2
infowars.com	0.28	257	0	0	0
faz.net	0.25	250	0	0	0
thegatewaypundit.com	0.13	226	0	0	0
whitehouse.gov	0.11	207	1,609	1,587	1,780
nu.nl	0.08	189	0	0	0
nos.nl	0.10	130	0	0	0
nzz.ch	0.08	110	0	0	0
wnd.com	0.45	109	0	0	0
dailycaller.com	0.40	107	0	0	0
yenisafak.com	0.13	95	0	0	0
azvision.az	0.10	92	0	0	0
arynews.tv	0.08	69	0	0	0
oilprice.com	0.09	69	0	0	0
trt.net.tr	0.09	61	0	0	0
theconservativetreehouse.com	0.12	42	0	0	0
bizpacreview.com	0.94	31	0	0	0
avaaz.org	0.13	28	0	0	0
bignewsnetwork.com	0.11	15	0	0	0
politico.mx	0.11	12	0	0	0
truepundit.com	0.19	12	0	0	0
lorientlejour.com	0.12	9	0	0	0
robinspost.com	2.51	8	0	0	0
conservativefiringline.com	0.33	7	0	0	0
ussanews.com	3.20	7	0	0	0
zazoom.it	0.42	3	0	0	0
newswithviews.com	0.08	2	0	0	0
bpr.org	0.08	1	0	0	0
conservativetribune.com	0.15	1	0	0	0
patriotretort.com	0.09	1	0	0	0
eaworldview.com	0.13	0	0	0	0
limportant.fr	0.11	0	0	0	0
mambolook.com	0.91	0	0	0	0
therebel.media	0.15	0	0	0	0
thetruthwins.com	0.12	0	0	0	0
truthfeednews.com	0.49	0	0	0	0
worthynews.com	0.16	0	0	0	0
48.pm	0.52	NA	0	0	0
back.ly	2.42	NA	0	0	0
bitchute.com	0.28	NA	0	0	0
breakingthenews.net	0.12	NA	0	0	0
caliberhitting.com	0.19	NA	0	0	0
commun.it	0.49	NA	0	0	0
cosiskey.com	0.28	NA	0	0	0
dmlnews.com	0.13	NA	0	0	0
dragplus.com	0.33	NA	0	0	0
floridahomeprepper.com	0.13	NA	0	0	0
gab.com	0.86	NA	0	0	0

(Continued on next page)

Domain	Partisanship	Reach	No. Occ. D	No. Occ. R	No. Occ. NP
geopoliting.com	0.14	NA	0	0	0
hayatsk.info	0.16	NA	0	0	0
hotpagenews.com	0.97	NA	0	0	0
ifilmfeatures.com	0.10	NA	0	0	0
inbound.li	0.08	NA	0	0	0
jenkers.com	0.56	NA	0	0	0
latinosfortrump.us	0.08	NA	0	0	0
lolsided.com	0.24	NA	0	0	0
magapill.com	1.56	NA	0	0	0
michaelsnyderforidaho.com	0.25	NA	0	0	0
nairapark.com	0.11	NA	0	0	0
nudesftw.com	0.55	NA	0	0	0
oddcrimes.com	0.82	NA	0	0	0
po.st	0.13	NA	0	0	0
prescient.info	0.24	NA	0	0	0
pscp.tv	2.06	NA	0	0	0
puppetstringnews.com	0.53	NA	0	0	0
rankeador.com.br	0.10	NA	0	0	0
rickwells.us	0.10	NA	0	0	0
rsbnetwork.com	0.09	NA	0	0	0
rviv.ly	0.11	NA	0	0	0
sco.lt	0.15	NA	0	0	0
shareasale.com	0.10	NA	0	0	0
spreaker.com	0.08	NA	0	0	0
storiesflow.com	0.61	NA	0	0	0
tacticalinvestor.com	0.43	NA	0	0	0
thegoldwater.com	0.59	NA	0	0	0
titrespresse.com	0.48	NA	0	0	0
trib.al	0.35	NA	0	0	0
trump-news.today	0.67	NA	0	0	0
vipscandals.com	1.60	NA	0	0	0
wi1848forward.blogspot.com	0.16	NA	0	0	0

Notes: The table lists all (canonical) domains randomly assigned to Republican users out of the top 100 most Republican domains identified based on the procedure described in Section B2 of the Online Appendix. The first column shows the domain names. The second column displays the corresponding value of the "Partisanship" measure, which indicates how partisan a domain identified as Republican is. This measure is based on the χ^2 test outlined in Section B1 of the Online Appendix. The corresponding χ_i^2 is a test statistic for the null hypothesis that the propensity to mention domain i in a tweet is equal for Democrats and Republicans (based on a sample of millions of tweets from Republican and Democratic Twitter users during the 2018 midterm elections). The higher the χ_i^2 value is, the more significantly a domain is primarily mentioned by supporters of one of the two parties. For all domains identified as highly partisan in this way, we look at whether they were more often used by Democrats or Republicans. In the table here, we only list domains that are identified as highly partisan and were more often used by Republicans, and we multiply the corresponding χ_i^2 value by $1,000$ for ease of readability. "Reach" refers to the reach (visitors per 1M web users) provided by AWIS (2020). Domains for which this data is missing in our dataset are indicated with NA. The remaining three columns indicate how often a domain occurs in users' election-related first search results pages during our study, separately by type of user (Democratic, Republican, non-partisan). The table is ordered by "Reach".

TABLE E5—RANDOM SELECTION OF 300 NON-PARTISAN SEARCH TERMS

Mali history	Burma culture	Grenadines history	736+414-4578-1268	968+8757	Korea sports
what are 5 miles in inches	what is 8 miles in mm	what is 4 mm in light years	wayfair	autozone	China GDP per capita
Spain sports	what are 1 ounces in tons	98/1316/7417	what is 3 kilometers in nautical miles	what is 1 tons in grams	what is 1 feet in light years
what are 2 liters in tsps	Romania culture	chase	what is 5 meters in inches	Ireland sports	Cameroon people
what is 5 tons in grams	spanish to english	182-168-3593	Panama GDP per capita	157/8673/8403-1683	673+5139/2308
Venezuela history	what are 10 ounces in tbsps	what is 5 miles in light years	what is 10 cm in mm	Togo neighbouring countries	Burma GDP per capita
Eritrea GDP	Russia population	what are 1 tbsps in ml	what is 4 tsps in ounces	Emirate people	what are 9 kilometers in cm
what are 8 kilometers in nautical miles	what are 10 tbsps in gallons	398/9379+8494	what are 8 tbsps in tsps	71+359	what is 5 kilos in pounds
airbnb	what are 1 tbsps in tsps	what is 7 gallons in ml	Malta history	scotland culture	yahoo mail
793+1728	what is 6 feet in nautical miles	407-1106	what are 5 kilos in grams	127/9489-6848+3673	St Lucia GDP
Saudi Arabia sports	San Marino sports	949-6932	325+9207+8139	963/529+451/9799	717/3824+9877
what are 2 grams in tons	595-7945/2597	what is 7 tsps in tbsps	what is 8 inches in mm	Argentina culture	Congo people
962/457-1930-7364	880-7263	Croatia GDP per capita	what is 4 nautical miles in feet	what is 3 grams in tons	Suriname sports
what is 9 meters in kilometers	Papua GDP	what is 1 ounces in gallons	what is 7 gallons in ounces	what is 5 kilos in grams	South Africa sports
634+9657	595-2846+8118	what are 8 ounces in tons	what is 1 meters in inches	Vanuatu population	what is 1 meters in miles
Ivory Coast GDP per capita	616+2128	what are 9 nautical miles in cm	Angola culture	in usa GDP	Syria GDP per capita
440/8117/3937	Kuwait population	487/8241+1238/6832	France sports	Tunisia GDP	Mongolia culture
amazon prime video	Sierra Leone neighbouring countries	what is 2 light years in feet	what are 9 nautical miles in miles	expedia	505+7251
what are 5 grams in ounces	what is 10 cm in nautical miles	636+9327	Korea people	East Timor neighbouring countries	224+8775+6307
what is 10 cm in miles	what is 5 tbsps in ounces	what are 1 pounds in grams	Italy population	Bosnia GDP per capita	694+4026+1933
Bosnia sports	839+7947	Herzegovina people	427-5105-6496-7384	339/1329	what are 2 ounces in liters
549/2764-6100	what is 2 ounces in ounces	Kyrgyzstan GDP	908-7084	what is 8 light years in feet	what is 4 light years in kilometers
Sierra Leone GDP per capita	487+3111-5904/431	Nicaragua GDP per capita	Namibia GDP	what are 5 kilometers in miles	Brunei GDP
St. Lucia population	Brunei GDP per capita	Myanmar population	Germany culture	what is 7 cm in kilometers	Turkmenistan culture
what are 4 grams in kilos	862-7857-7173	what is 1 inches in light years	what are 8 mm in cm	Singapore culture	what are 5 tbsps in gallons
what is 9 ml in tbsps	what is 1 ml in liters	what is 10 tsps in ml	55/9690+5373+2591	what are 3 ounces in liters	Antigua people
what is 5 ml in tbsps	379-2502	Finland population	20+8741-5644+6141	Malawi GDP	498/755
524/4132	what are 1 light years in kilometers	what are 9 ounces in pounds	what are 5 light years in cm	what are 4 miles in cm	what is 2 ounces in pounds
what is 10 kilometers in feet	587+770	131+6529-7843	880-622/8769-3602	what is 3 inches in kilometers	Switzerland population
what is 8 tbsps in ml	what are 9 ounces in gallons	78+6018	140-1923+1081	453/1144	582+5797/4975
735/7898+3425	what is 6 liters in gallons	what is 10 tbsps in ml	Gabon GDP	what are 10 tons in grams	what is 5 ounces in liters
what are 1 ounces in tbsps	what is 8 grams in ounces	what are 6 pounds in grams	Croatia people	815+481+4762+9250	Zambia history
Poland sports	what is 6 light years in cm	Grenadines neighbouring countries	Nauru history	962-4590/82-3815	what are 5 light years in mm
Saudi Arabia GDP per capita	what is 3 kilometers in light years	533+2938	Hungary history	wales GDP	Azerbaijan GDP
958/8459/5292/4805	217-9697-9653-978	what is 7 gallons in tbsps	what is 8 pounds in grams	what is 2 kilometers in inches	447+1397
411/4287	Greece neighbouring countries	16/1412/155-7189	547+4359/7081	what is 6 tbsps in gallons	Panama people
967/5689/9413	186+6700-3860	529/8619-9683+375	Luxembourg sports	what is 9 tbsps in tsps	Madagascar population
what is 9 ounces in gallons	Cambodia people	what are 7 meters in kilometers	Uzbekistan people	what are 4 meters in miles	what is 2 inches in nautical miles
what is 10 tbsps in gallons	what are 4 kilos in ounces	Azerbaijan sports	Seychelles GDP per capita	Poland population	Burundi people
what is 7 tsps in gallons	Tajikistan population	what are 2 ml in tsps	Bahrain GDP	what are 6 ml in tsps	Marshall population
252/5687+8472	what is 8 mm in meters	863/2198	nfl	what are 1 inches in mm	Santa Lucia people
what is 6 ounces in pounds	what is 8 pounds in tons	Burma neighbouring countries	Lithuania people	xfinity	what is 3 tbsps in liters
374+9718+6631	486/9557-1652+9663	Gambia neighbouring countries	54+9585+6903+2311	541-9729/4967/3974	what is 9 miles in nautical miles
Jamaica culture	what is 6 tbsps in tsps	what is 2 mm in meters	514+818+608	what are 10 pounds in ounces	what are 1 ml in gallons
Somalia sports	Bhutan GDP per capita	856/6664+3085	what are 2 kilometers in inches	what is 5 ounces in grams	285-3933/6300
Grenadines GDP per capita	what are 6 ml in gallons	523-1434-422	what is 5 ml in ounces	what are 10 feet in kilometers	631/1342-2800+994
746-5249-163+1067	149+7018/6373	what are 8 kilos in ounces	706/822	Rwanda history	Albania sports
what are 5 gallons in liters	Kazakhstan neighbouring countries	Rwanda sports	Maldives history	St. Kitts neighbouring countries	Libya sports
Syria neighbouring countries	Belize culture	Belarus GDP per capita	1-2703/3031/41	Pakistan history	526/9087/4738
Bolivia population	wales GDP per capita	Gambia population	what is 2 light years in miles	529/4531	what is 9 inches in miles

Notes: The table lists a random selection of 300 non-partisan search terms that were used at least once by one of the bots in our study. Overall, a total of 2,581 non-partisan search terms were used during our study. Section B3 in the Online Appendix provides a detailed account of the generation, selection, and assignment of these terms.

TABLE E6—PARTISAN SEARCH TERMS ASSIGNED TO DEMOCRATIC USERS

#bernie	#bidenharris	#blm	#communism
#democracy	#dumptrump	#fakenews	#feelthebern
#fucktrump	#government	#impeachtrump	#kamalaharris
#liberals	#notmypresident	#obama	#political
#progressive	#resist	#socialism	#voteblue
address gun	affordable care	affordable care act	affordable health
affordable insurance	affordable quality care	african american	air and water
air to water	air water	all things must pass	american community survey
attack women	big oil	birth control	birth control pills
black african american	black and brown	black white supremacist	black women
call donald trump	call trump	care act	care coverage
center for reproductive health	change real	check act	children and family
children parents	choose the right	civil rights	civil rights act
civil rights act passed	civil rights movement	climate change	community of color
congressional republican	criminal background check	detention center	donald trump
enrollment health	environmental protection	environmental protection agency	equal pay
equal pay act	equal rights	equal rights amendment	every dollar app
existing conditions	fair labor act	fair pay act	fair wage
fair wage act	federal minimum wage	fight at work	fight for justice
fight justice	fight work	for profit	for profit college
free background check	gender disorder	gender identity	gender identity disorder
gender pay gap	gift tax	girls to women	gun safety
gun safety course	health care	health care coverage	health coverage
health insurance	health insurance affordable	health insurance coverage	heart
heart attack women	hiv aids	hiv and aids	house gop
housing	how to	how to fight	insurance
insurance coverage	internet explorer	intimate partner	juvenile
juvenile detention	law of one	let go	living wage
loan debt	lose health	lose weight	low income
low income housing	make ends meet	make health	medicare and medicaid
medicare enrollment	medicare enrollment period	medicare medicaid	million dollar women
million women	minimum living wage	minimum wage	moms demand action
must fight	must pass	my access	one law
one million	one million women	open enrollment	open internet
paid sick leave	passed act	pay equal	pay gap
pay wall	people color	people of color	policy on immigration
pre existing	profit college	protect democracy	protected health
protected health information	protection workers	quality affordable	raising minimum
raising minimum wage	reproductive health	reproductive health services	reproductive rights
republican congressional committee	rights act	rights of women	rights of workers
sandy hook	school shooting	senate gop	separation children
separation family	sexual orientation	sign health	stimulus check
student debt	student loan	student loan debt	the dream
the dream act	the real	the right woman	the un american
trump admin	trump gop	trump want	turn back
turn back time	united healthcare	united healthcare coverage	united to protect democracy
us immigration	us immigration policy	violence women	what are working conditions
what is minimum wage	what sexual orientation	white supremacist	white supremacist trump
will lose	woman right	women access	women against violence
women color	women family	women girls	women in black
women of color	women rights	womens health center	work and family
work family	workers protection act	workers rights	working conditions
you will lose	zero tolerance		

Notes: The table lists the partisan search terms assigned to Democratic users, consisting of the top partisan hashtags and the slanted search terms compiled based on the procedure outlined in Section B4 of the Online Appendix.

TABLE E7—PARTISAN SEARCH TERMS ASSIGNED TO REPUBLICAN USERS

#alllivesmatter	#american	#bluelivesmatter	#buildthewall
#conservativememes	#conservatives	#coronavirus	#draintheswamp
#foxnews	#patriot	#potus	#presidenttrump
#progun	#republicanmemes	#republicanparty	#trumpsupporters
#walkaway	#wga	#whitehouse	#wwg
abortion	america	american people	canada mexico
cut jobs	god bless	god bless america	government control
great news	illegal immigrants	illegal immigration	immigration
jobs act	men and women	men women	mexico to canada
nancy pelosi	nancy pelosi trump	one plus one	one size
plus size	pro abortion	pro choice	pro life
pro life abortion	pro life pro choice	size fit	tax code
the american people	the cut	the government	the great
the jobs act	the unemployment rate	tire size	trump
unemployment rate	women for men	women in uniform	women uniform

Notes: The table lists the partisan search terms assigned to Republican users, consisting of the top partisan hashtags and the slanted search terms compiled based on the procedure outlined in Section B4 of the Online Appendix.

TABLE E8—SEARCH TERMS USED IN ELECTION-RELATED GOOGLE SEARCHES

Joe Biden
Republican
Democrat
White House
Donald Trump
Mail-in ballot
Congress
Polling station
voter fraud
Arizona
stop count
count votes
illegal ballots
Georgia
Pennsylvania
Michigan
electoral vote
Wisconsin
google news
senate race
presidential transition
did my vote count?
Joe Biden is
Georgia recount
Michigan election
election results 2020
has pennsylvania certified the election
covid vaccine
mask mandate
has michigan certified the election
curfew
Donald Trump is
Sidney Powell
trump supporters dc
senate results georgia
capitol washington dc
hyde smith
president elect
trump supporters
who won senate 2021
election certification
riot capitol hill
house impeachment
capitol police officer dies
joint chiefs of staff letter
inaguration day threats
caldwell oath keepers
inaguration day 2021
presidential pardons list 2021
national guard in capitol
capitol rioters
Biden on student loans
Biden on immigration
Biden on gun control
football
Biden on stimulus
Biden executive orders
Biden news

Notes: The table lists all 58 election-related search terms used in our study ordered according to the date they were selected/used. The terms were selected from Google Trends (2020, 2021) feature pages focusing on the most frequently used searches on the US 2020 election and closely related topics such as the presidential transition and Biden's inauguration. Section B6 of the Online Appendix provides additional details on the compilation of these terms from Google Trends. The term "football" may refer to the so-called nuclear football, which Trump did not hand over to Biden in person (see, e.g., `https://edition.cnn.com/2021/01/19/politics/trump-biden-nuclear-football-inauguration/index.html` or `https://www.foxnews.com/politics/nuclear-football-handoff-between-trump-biden`).

TABLE E9—TOP-50 DOMAINS IN SEARCH RESULTS BY CATEGORY

National news	Government/elections	Information platforms	Local news	Business domains	Other domains
cnn.com	cdc.gov	en.wikipedia.org	ajc.com	play.google.com	bbc.com
nytimes.com	whitehouse.gov	facebook.com	latimes.com	biography.com	theguardian.com
washingtonpost.com	ballotpedia.org	twitter.com	observer.com	amazon.com	brookings.edu
politico.com	joebiden.com	britannica.com	freep.com	pfizer.com	brennancenter.org
nbcnews.com	justice.gov	history.com	detroitnews.com	apps.apple.com	heritage.org
npr.org	gsa.gov	support.google.com	cincinnati.com	aranzulla.it	reuters.com
apnews.com	michigan.gov	youtube.com	inquirer.com	marketwatch.com	aarp.org
abcnews.go.com	history.house.gov	who.int	mlive.com	snopes.com	whitehousehistory.org
cbsnews.com	congress.gov	dictionary.com	clickondetroit.com	harpersbazaar.com	raps.org
usatoday.com	loc.gov	merriam-webster.com	bridgemi.com	dailywritingtips.com	visitthecapitol.gov
thehill.com	fda.gov	ydr.com	clickorlando.com	criminal.findlaw.com	politifact.com
foxnews.com	usa.gov	fr.wikipedia.org	nymag.com	kiplinger.com	ncsl.org
businessinsider.com	sos.ca.gov	blog.twitter.com	pennlive.com	techcrunch.com	mayoclinic.org
usnews.com	covid19.ca.gov	factcheck.org	abc7news.com	daveramsey.com	military.com
cnbc.com	coronavirus.wa.gov	govtrack.us	texastribune.org	getbellhops.com	statnews.com
axios.com	presidentialtransition.org	classroommagazines.scholastic.com	nbcwashington.com	nerdwallet.com	nature.com
nypost.com	senate.gov	it.wikipedia.org	6abc.com	efile.com	channelnewsasia.com
vox.com	indy.gov	en.m.wikipedia.org	abc11.com	vexillology.fandom.com	upenn.edu
militarytimes.com	washington.org	news.google.com	khou.com	business-standard.com	uk.usembassy.gov
thedailybeast.com	vote.org	refinery29.com	fox5dc.com	marieclaire.com	the-sun.com
pbs.org	roy.house.gov	cookpolitical.com	webmd.com	nationsonline.org	independent.co.uk
bloomberg.com	aoc.gov	asu.edu	post-gazette.com	bankrate.com	arizona.edu
forbes.com	270towin.com	avalara.com	seattletimes.com	nomadicmatt.com	gsu.edu
projects.fivethirtyeight.com	federalappeals.com	whyy.org	wusa9.com	studentloanhero.com	uga.edu
wsj.com	archives.gov	blog.google	katc.com	astrazeneca.com	nejm.org
newyorker.com	uscp.gov	es.wikipedia.org	news4jax.com	healthinsurance.org	georgia.org
cnet.com	ncsbe.gov	onlyinyourstate.com	apps.bostonglobe.com	matadornetwork.com	opensecrets.org
news.bloomberglaw.com	health.state.mn.us	contagionlive.com	miamiherald.com	ziprecruiter.com	technologyreview.com
huffpost.com	media.pa.gov	tripsavvy.com	boston.com	coolkidfacts.com	govexec.com
time.com	donaldjtrump.com	apmresearchlab.org	abc7.com	artsandculture.google.com	psu.edu
newsnationnow.com	browardsoe.org	news.google.it	deseret.com	michiganbusiness.org	espn.com
edition.cnn.com	healthaffairs.org	news.google.ru	washingtonian.com	safehome.org	visitpa.com
theatlantic.com	house.gov	news.google.fr	wnep.com	nrn.com	spacepolicyonline.com
int.nyt.com	fec.gov	wilsoncenter.org	michiganradio.org	azbigmedia.com	nfl.com
reason.com	sos.state.mn.us	rgj.com	wavy.com	hicapitoldc.com	niemanlab.org
motherjones.com	defense.gov	collinsdictionary.com	fairfieldsuntimes.com	insidehighered.com	defenseone.com
foreignpolicy.com	fbi.gov	bestplaces.net	kusi.com	ecode360.com	dw.com
buzzfeednews.com	health.pa.gov	sites.google.com	sfchronicle.com	smartasset.com	cato.org
cdn.cnn.com	georgia.gov	alarms.org	whio.com	mapsofworld.com	health.clevelandclinic.org
thenation.com	obamawhitehouse.archives.gov	de.wikipedia.org	abc7chicago.com	morethanmindgames.com	sciencemag.org
vogue.com	electoral-vote.com	thesaurus.yourdictionary.com	wlbt.com	sf.eater.com	telegraph.co.uk
newsweek.com	federalregister.gov	it.m.wikipedia.org	mynbc15.com	seattle.eater.com	theconversation.com
ap.org	jcs.mil	dictionary.cambridge.org	newstimes.com	livability.com	thecrazytourist.com
c-span.org	elections.maryland.gov	google.com	thenevadaindependent.com	homesnacks.com	youthrights.org
apps.washingtonpost.com	accd.vermont.gov	capradio.org	azcentral.com	whitehouseblackmarket.com	michigan.org
nybooks.com	azdhs.gov	simple.wikipedia.org	cbssports.com	mololamken.com	umich.edu
people.com	headcount.org	apartmentlist.com	lansingstatejournal.com	livescience.com	msu.edu
salon.com	pa.gov	vocabulary.com	wesh.com	chrome.google.com	healthline.com
atlanticcouncil.org	gop.com	oldest.org	kcra.com	maxifoot.fr	uchealth.org
armytimes.com	sos.state.co.us	ourworldindata.org	wtop.com	deadline.com	ourpublicservice.org

Notes: The table lists the 50 domains most often occurring in election-related search results by category. Domains in each category are ordered according to their number of occurrences (from most to least occurring).

TABLE E10—MAIN RESULTS AND ROBUSTNESS CHECKS ON BEHAVIOR-BASED PERSONALIZATION

	Dependent variable:							
	No. preferred domains in search results						P(any preferred)	Top rank
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$51-100$ visits	0.154							
	(0.072)							
$101-150$ visits	0.432							
	(0.142)							
$151+$ visits	0.560							
	(0.164)							
# visits (in tens)		0.036	0.067	0.050	0.048	0.033	0.012	−0.034
		(0.014)	(0.018)	(0.019)	(0.018)	(0.016)	(0.005)	(0.030)
Mean Dependent Variable	2.06	2.06	2.06	1.95	1.95	2.29	0.85	4.16
User FE	X		X	X	X	X	X	X
Date of search FE	X	X	X	X	X	X	X	X
Search term FE					X	X	X	X
Clust. SE user	X	X	X	X	X	X	X	X
Clust. SE date	X	X	X	X	X	X	X	X
Clust. SE search term					X	X	X	X
Observations	16,142	16,142	16,142	24,987	24,987	21,268	24,987	21,268
R^2	0.408	0.272	0.409	0.351	0.433	0.398	0.237	0.479

Notes: This table presents the regression output displayed in Figure 2C in column (1) (see the notes to Figure 2 for details) and various robustness tests in columns (2)–(8). The units of observation are user × day in columns (1)–(3) and user × election-related search in columns (4)–(8). The sample is restricted to observations where users see at least one preferred domain in the first search results page in columns (6) and (8). The dependent variable is the number of preferred domains in the first search results page for election-related queries in columns (1)–(6), an indicator variable equal to 1 if this page contains any of the user's preferred domains in column (7), and the rank of the highest-ranking domain preferred by the user in this page (with smaller values indicating higher placement in the search results) in column (8). #*visits* measures the total number of times a user has previously visited its preferred domains. Fixed effects and standard error clustering levels are indicated for each specification.

TABLE E11—MAIN RESULTS AND ROBUSTNESS CHECKS ON LOCATION-BASED PERSONALIZATION

	Dependent variable:								
	$P(Domaininresults)$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1 other user in city sees domain	0.287 (0.006)								
2 other users in city see domain	0.436 (0.010)								
3 other users in city see domain	0.581 (0.015)								
4 other users in city see domain	0.732 (0.015)								
5 other users in city see domain	0.888 (0.017)								
# other users in city see domain		0.173 (0.004)	0.170 (0.006)	0.183 (0.005)	0.166 (0.007)				
# other users in state see domain						0.008 (0.001)	0.007 (0.002)	0.008 (0.002)	0.008 (0.002)
# users overall see domain	0.001 (0.0001)	0.001 (0.0002)	0.001 (0.0003)	0.001 (0.0002)	0.002 (0.0003)	0.006 (0.0001)	0.006 (0.0002)	0.006 (0.0002)	0.006 (0.0002)
Sample	all	all	D	R	nonpartisan	all	D	R	nonpartisan
Observations	15,343,800	15,343,800	5,114,600	5,114,600	5,114,600	15,343,800	5,114,600	5,114,600	5,114,600
R^2	0.715	0.711	0.701	0.714	0.719	0.642	0.633	0.636	0.658

Notes: This table presents the regression output displayed in Figure 2D in column (1) (see the notes to Figure 2 for details) and various robustness tests in columns (2)–(9). The units of observation are user $\times$ domains in election-related search results. The sample is restricted to observations from Democratic users in columns (3) and (7), Republican users in columns (4) and (8), and non-partisan users in columns (5) and (9). The dependent variable is an indicator variable equal to 1 if a specific domain occurs on a user's first search results page. Explanatory variables indicate the number of other users in the same city or the same state who see the given domain. All columns control for the number of overall users who see the given domain and include user, date of search and domain fixed effects. Standard errors are three-way clustered by user, date of search and domain.

TABLE E12—FURTHER ROBUSTNESS CHECKS ON LOCATION-BASED PERSONALIZATION

	$P(Domaininresults)$					
	(1)	(2)	(3)	(4)	(5)	(6)
Majority of users in city see domain	0.387	0.390	1.826			
	(0.016)	(0.017)	(0.015)			
# users overall see domain	0.004	0.004	0.067	0.004	0.004	0.070
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Majority of users in state see domain				0.332	0.335	1.157
				(0.016)	(0.017)	(0.015)
Estimation Method	OLS	OLS	Logit	OLS	OLS	Logit
User FE	X	X	X	X	X	X
Date of search FE	X	X	X	X	X	X
Domain FE		X			X	
Clust. SE user	X	X	X	X	X	X
Clust. SE date	X	X	X	X	X	X
Clust. SE domain		X			X	
Number of Observations	15,343,800	15,343,800	15,343,800	15,343,800	15,343,800	15,343,800
R^2	0.673	0.673		0.662	0.662	

Notes: This table presents further robustness test related to Figure 2D. The units of observation are user × domains in election-related search results. The dependent variable is an indicator variable equal to 1 if a specific domain occurs on a user's first search results page. Explanatory variables are indicator variables that are equal to 1 if the majority of the other users in the same city or the same state see the given domain. All columns control for the number of overall users who see the given domain. Fixed effects and standard error clustering levels are indicated for each specification. Columns (3) and (6) use a logistic rather than a linear regression model. The estimation of the logistic regression with fixed effects is implemented based on the approach suggested by Stammann (2018).

TABLE E13—MAIN RESULTS AND ROBUSTNESS TESTS ON THE EFFECTS OF LOCATION AND PARTISANSHIP ON SEARCH RESULTS IDEOLOGY

	Dependent variable: search results ideology score												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
D user	0.042	−0.575	0.016	−0.704	−0.114	−0.370	0.022	−0.425	−0.094	−0.229	0.107	−0.296	−0.051
	(0.211)	(0.127)	(0.406)	(0.259)	(0.054)	(0.077)	(0.333)	(0.114)	(0.092)	(0.170)	(0.087)	(0.241)	(0.045)
R user	0.139	−0.268	0.107	−0.430	−0.055	−0.270	0.153	−0.190	−0.101	−0.299	−0.032	−0.223	−0.010
	(0.226)	(0.262)	(0.508)	(0.209)	(0.125)	(0.127)	(0.267)	(0.205)	(0.084)	(0.065)	(0.094)	(0.041)	(0.060)
D city	−0.525	−0.591	−2.335	0.263	−0.350	−0.511	−1.906	0.290	−0.115	0.106	0.201	0.054	−0.495
	(0.302)	(0.387)	(0.774)	(0.343)	(0.407)	(0.465)	(1.385)	(0.258)	(0.178)	(0.205)	(0.281)	(0.208)	(0.312)
R city	0.435	1.182	1.973	0.855	0.524	0.954	1.814	0.665	0.130	0.266	0.423	0.296	0.366
	(0.230)	(0.369)	(0.687)	(0.447)	(0.191)	(0.477)	(0.969)	(0.281)	(0.104)	(0.338)	(0.211)	(0.204)	(0.171)
Sample	All	All news	Local news	National news	All	All news	Local news	National news	All	All	All	All	All
Unit of obs.	User/day	User/day	User/day	User/day	User/SERP	User/SERP	User/SERP	User/SERP	User/SERP	User/SERP	User/SERP	User/SERP	User/SERP
Ideology measure	Joint	Joint	Joint	Joint	Joint	Joint	Joint	Joint	Bakshy et al.	Budak	Mturk	Pew	Robertson et al.
Mean dependent variable	-36.94	-69.41	-14.06	-77.49	-36.71	-70.7	-14.61	-79.28	-24.98	-55.47	-15.73	-59.51	-17.79
% domains in index	84.56	98.17	95.65	98.98	84.56	98.17	95.65	98.98	50.84	22.59	58.51	31.47	84.56
D - R user = 0 (χ^2)	0.10 (0.20)	0.31 (1.17)	0.09 (0.05)	0.27 (0.99)	0.06 (0.17)	0.10 (0.56)	0.13 (0.38)	0.24 (1.34)	0.01 (0.02)	0.07 (0.11)	0.14 (2.09)	0.07 (0.10)	0.04 (0.30)
D - R city = 0 (χ^2)	0.96 (6.82)	1.77 (9.42)	4.31 (12.10)	0.59 (1.33)	0.87 (3.10)	1.47 (2.88)	3.72 (2.52)	0.37 (1.18)	0.24 (7.79)	0.16 (0.72)	0.22 (1.22)	0.24 (52.93)	0.86 (3.49)
Date of search FE	X	X	X	X	X	X	X	X	X	X	X	X	X
Language FE	X	X	X	X	X	X	X	X	X	X	X	X	X
Search term FE					X	X	X	X	X	X	X	X	X
Clust. SE user	X	X	X	X	X	X	X	X	X	X	X	X	X
Clust. SE date	X	X	X	X	X	X	X	X	X	X	X	X	X
Clust. SE search term					X	X	X	X	X	X	X	X	X
Observations	13,091	12,889	7,378	12,580	24,987	22,999	9,570	22,024	24,987	24,987	24,987	24,987	24,987
R^2	0.838	0.667	0.745	0.617	0.849	0.669	0.823	0.641	0.573	0.594	0.738	0.651	0.786

Notes: This table presents the regression outputs displayed in Figure 3 in column (1)–(4) (see the notes to Figure 3 for details) and various robustness tests in columns (5)–(13). The units of observation are user × day in columns (1)–(4) and user × election-related search results in columns (5)–(13). The dependent variable is the Search Result Ideology Score (SRIS, see equation (2) in the main text) for different sets of domains listed in the first search results pages to election-related queries: All domains in columns (1), (4), and (9)–(13), all news domains in columns (2) and (6), local news domains in column (3) and (7), and national news domains in columns (4) and (8). The SRIS is computed based on the five different domain ideology indices in columns (1)–(8) and on a single domain ideology index indicated in the row "Ideology measure" in columns (9)–(13). The row "% domains in index" indicates the share of the respective domains covered by the corresponding indices. The SRIS is measured on a liberal-conservative scale of [-100,100] in all cases. *D user* (*R user*) is an indicator variable equal to 1 if the user has been assigned Democratic (Republican) browsing and search preferences (with non-partisan users as reference category). *D city* (*R city*) is an indicator variable equal to 1 if the user has been assigned to a city with predominantly Democratic (Republican) voters (with purple cities as reference category). Fixed effects and standard error clustering levels are indicated for each specification. The row "D - R user = 0 (χ^2)" presents the absolute difference between the coefficient estimates for the Democratic and Republican user indicators, as well as the chi-squared value (in parenthesis) of the linear hypothesis test of both coefficients being of equal size. The row "D - R city = 0 (χ^2)" shows the same for the Democratic and Republican city indicators.

TABLE E14—FURTHER ROBUSTNESS TESTS ON THE EFFECTS OF LOCATION AND PARTISANSHIP

	Dependent variable: search results ideology score			
	(1)	(2)	(3)	(4)
Visited sites ideology	0.003	0.012	−0.008	0.005
	(0.007)	(0.011)	(0.016)	(0.014)
Share Rep. voters in city	0.024	0.043	0.114	0.012
	(0.009)	(0.014)	(0.037)	(0.012)
Sample	All	News domains	Local news	National news
Observations	24,987	22,999	9,570	22,024
R^2	0.562	0.422	0.679	0.410

Notes: This table presents robustness tests related to Figure 3 using more fine-grained measures of user and city ideology. The units of observation are user $\times$ election-related search results. The dependent variable is the Search Result Ideology Score (SRIS, see equation (2) in the main text) for different sets of domains listed in the first search results pages to election-related queries: All domains in columns (1), all news domains in columns (2), local news domains in column (3), and national news domains in columns (4). *Visited sites ideology* is the average ideology score of the preferred domains visited by the user. *Share Rep. voters in city* is the percentage share of Republican voters in the 2016 US election in the city where the user is located. All columns include date of search and browser language fixed effects. Standard errors are two-way clustered by user and date of search.

REFERENCES

Ahrefs. 2020. "Top 100 Most Visited Websites in the US." https://ahrefs.com/blog/most-visited-websites (accessed: October 1, 2020).

AWIS. 2020. "Alexa Web Information Service." https://aws.amazon.com/awis (accessed: June 1–15, 2020).

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science*, 348(6239): 1130–1132.

Best Hashtags. 2020*a*. "Best Democrat Hashtags." http://best-hashtags.com/hashtag/democrat (accessed: October 20, 2020).

Best Hashtags. 2020*b*. "Best Republican Hashtags." http://best-hashtags.com/hashtag/republican (accessed: October 20, 2020).

Budak, Ceren, Sharad Goel, and Justin M. Rao. 2016. "Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis." *Public Opinion Quarterly*, 80(S1): 250–271.

Dottle, Rachel. 2019. "Where Democrats and Republicans Live in your City." https://projects.fivethirtyeight.com/republicans-democrats-cities (accessed: October 20, 2020).

Federal Election Commission. 2022. "Federal Elections 2020." https://www.fec.gov/introduction-campaign-finance/election-results-and-voting-information/federal-elections-2020 (accessed: June 2022).

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica*, 78(1): 35–71.

Gentzkow, Matthew, Jesse Shapiro, and Matt Taddy. 2019. "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech." *Econometrica*, 87(4): 1307–1340.

Google Trends. 2020. "Google Trends Data." https://trends.google.com (accessed: October 20 to December 31, 2020).

Google Trends. 2021. "Google Trends Data." https://trends.google.com (accessed: January 1 to February 5, 2021).

Matter, Ulrich, and Roland Hodler. 2024. "Data and Code for: Web Search Personalization During the US 2020 Election." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E1209061V1.

Mitchell, Amy, Jeffrey Gottfried, Jocelyn Kiley, and Katerina E. Matsa. 2014. "Political Polarization & Media Habits." Pew Research Center.

NewsGuard. 2023. "NewsGuard Ratings and Data." https://newsguardtech.com (accessed: April 24, 2023).

Peterson, Erik, and Shanto Iyengar. 2021. "Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?" *American Journal of Political Science*, 65(1): 133–147.

Poole, Keith T., and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science*, 29(2): 357–384.

Robertson, Ronald E., Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. 2018. "Auditing Partisan Audience Bias within Google Search." *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW): 148.

Stammann, Amrei. 2018. "Fast and Feasible Estimation of Generalized Linear Models with High-Dimensional k-way Fixed Effects." arXiv Preprint 1707.01815.

Wrubel, Laura, Justin Littman, and Dan Kerchner. 2019. "2018 U.S. Congressional Election Tweet Ids." Harvard Dataverse. https://doi.org/10.7910/DVN/AEZPLU.