It matters how you google it? Using agent-based testing to assess the impact of user choices in search queries and algorithmic personalization on political Google Search results

van Hoof, M.; Trilling, D.; Moeller, Judith; Meppelink, C.S.

DOI
10.1093/jcmc/zmae020

Publication date
2024

Document Version
Final published version

Published in
Journal of Computer-Mediated Communication

Citation for published version (APA):
van Hoof, M., Trilling, D., Moeller, J., & Meppelink, C. S. (2024). It matters how you google it? Using agent-based testing to assess the impact of user choices in search queries and algorithmic personalization on political Google Search results. *Journal of Computer-Mediated Communication*, *29*(6), Article zmae020. https://doi.org/10.1093/jcmc/zmae020

Journal of Computer-Mediated Communication, 2024, **29(6)**, zmae020
https://doi.org/10.1093/jcmc/zmae020
Full-length Research Article

International Communication Association

OXFORD

It matters how you google it? Using agent-based testing to assess the impact of user choices in search queries and algorithmic personalization on political Google Search results

Marieke van Hoof [1,*], Damian Trilling [2], Judith Moeller [3], Corine S. Meppelink [1]

[1]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, The Netherlands
[2]Department of Language, Literature and Communication, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[3]University of Hamburg—Leibniz Institute for Media Research | Hans Bredow Institute, Hamburg, Germany
*Corresponding author: Marieke van Hoof. Email: m.vanhoof@uva.nl

Abstract

Search engines, as key sources of political information, have sparked concerns regarding selective exposure driven by user choices based on political beliefs and filter bubbles created by algorithms. This study focuses on the most influential yet often-overlooked user choice: search queries. We investigate the extent to which user choices (search queries) and algorithmic personalization (search history) lead to divergent search results. Building on research linking immigration- and climate-related search queries to political characteristics, we conduct an experiment on Google Search employing agent-based testing. Using computational methods, we examine variations in search results by source (type) and search result page features. Our findings show that it is the specific search queries made by users, not algorithmic personalization, that lead to substantially divergent information sources in search results. This suggests future research should prioritize user choices in information search rather than control for them.

Lay Summary

Search engines like Google influence what political information we come across online. Many are concerned that the choices we make based on our political beliefs and the way Google's algorithms tailor search results to our past queries may limit the diversity of political information we encounter. This study focuses on a very influential but often-overlooked aspect of user choice: the keywords we type into the search bar. We investigate how these choices and algorithms lead to different search results. Building on earlier research findings, we set up an experiment that mimics how people search for information about immigration and climate change on Google Search. We look into how these factors can lead to different sources (types) and ways information is presented on the search result page. We find that the specific search queries we choose, not algorithms, lead to divergent information sources on the search result page. This finding suggests that if we want to understand how people are exposed to political information via search engines, we should pay more attention to the choices they make when searching for information, rather than just focusing on the technology behind search engines.

Keywords: search engines, political information, search queries, user choice, algorithmic personalization.

In today's high choice media environment, the public increasingly relies on algorithmic tools to navigate political information and news, with search engines serving as primary pathways to political information (Wojcieszak et al., 2022). There has been a notable focus on how search engines deliver personalized political information and news, through both *user choices* and *algorithmic personalization*.

Search engines incorporate mechanisms for users to affect the curation of search results via intentional choices. This corresponds to a large field of existing research on political selective exposure that shows that individuals that are given control over their information environment are driven by their pre-existing political beliefs and interests in their selection of political content (Stroud, 2008). Additionally, search engine algorithms filter and rank information based on observed search behavior, tailoring search results more to users' (inferred) preferences and interests.

A prominent concern in academic and public discourse regarding these mechanisms is their potential to restrict exposure to diverse political content through the creation of highly individualized information diets, or so-called "echo chambers" (Sunstein, 2001) and "filter bubbles" (Pariser, 2011). Whether driven by user choices or algorithmic personalization, the exposure to divergent information on the same political issues may result in a lack of a common ground of information. This challenges the ideal of a deliberate democratic society, which depends on a common ground for meaningful dialogue among individuals with different ideas and opinions (Helberger, 2019).

While concerns about filter bubbles persist, existing research generally does *not* support the notion that search engines lead to a narrowing of political information exposure (Robertson et al., 2018) or consumption (Courtois et al., 2018). One explanation could be that this research typically does not consider the pivotal role of user choices in shaping information search. While recent research underscores the importance of choices made by users rather than algorithms in the selection of news sources (Robertson et al., 2023), we focus on the primary factor that empowers user choices in information search: the formulation of search queries, which

Associate Editor: Cuihua Shen
Received: March 7, 2024. **Revised:** July 20, 2024. **Accepted:** September 6, 2024

marks the beginning of any search session. Although a couple of recent studies have addressed individual choices in political search queries (Blassnig et al., 2023; Menchen-Trevino et al., 2023; van Hoof et al., 2024), we know little about how these choices affect search result curation or how their impact compares to algorithmic personalization.

Therefore, this article aims to investigate the extent to which user choices in search queries and algorithmic personalization lead to divergent Google Search results. We examine these effects for two contentious, political issues in the Dutch context: immigration, a long-standing and polarizing issue (Boomgaarden & Vliegenthart, 2007), and climate change, a more recent and less polarizing issue (Wonneberger et al., 2020). Additionally, we provide insights into the variations in terms of (types of) sources and Google Search Engine Result Pages (SERPs) features (e.g., Top Stories, Knowledge Panel).

Given the opaque nature of digital platform algorithms, isolating the effects of user choices in search queries from algorithmic personalization poses a challenge. To address this, we employ agent-based testing (ABT; Haim, 2020) to collect SERPs by creating virtual agents that emulate human search behavior under systematically varied conditions. The research design builds on survey research by van Hoof et al. (2024), who identified clusters of individuals using distinct search queries.

Theoretical framework and related research

Search engines act as "algorithmic gatekeepers" by selecting and prioritizing the vast amount of information available to individuals seeking information online (Diakopoulos et al., 2018). To this end, search engines use various factors, such as geolocation, website popularity and freshness, to determine the most "relevant" search results (Google, 2023). Over the past decade, academic attention has been specifically directed toward two methods through which digital media offer personalized experiences: explicit and implicit personalization (Thurman and Schifferes 2012; also see Zuiderveen Borgesius et al., 2016).

Search engines afford so-called explicit personalization by allowing their users to make deliberate choices that affect the curation (i.e., selection and ranking) of search results, such as selecting search results or entering search queries. Additionally, search engines curate search results based on a user's past online behavior, which is used to infer their preferences and deliver more relevant content, so-called implicit personalization (Thurman & Schifferes, 2012). Google Search has incorporated mechanisms for inferring user preferences since 2009,[1] such as users' prior search queries and selected search results (Google, 2023). Importantly, this form of personalization provides content to users without their active consent (Zuiderveen Borgesius et al., 2016) and thus reduces their control and agency over search results. In this work, we refer to these two personalization methods such as *user choices in search queries* and *algorithmic personalization*, respectively.

Algorithmic personalization

Responding to public concerns about "filter bubbles" (Pariser, 2011), a series of studies in the past decade have investigated how search engine algorithms deliver personalized political information, potentially limiting the diversity of information consumed by the public. Despite these worries, survey research indicates that search engines can facilitate the discovery of news sources beyond one's regular routine. Studies relying on digital trace data support these findings by investigating the impact of different access points to political information. Findings indicate that search engines drive their users to consume more ideologically balanced or dissimilar information sources (Cardenal et al., 2019; Flaxman et al., 2016; Fletcher & Nielsen, 2018; Wojcieszak et al., 2022). Yet, while these studies draw conclusions about the impact of algorithmic personalization, the extent to which these outcomes can be attributed to algorithmic decision-making remains unclear due to their focus on consumption patterns.

To address this question, a series of search engine algorithm audits have been conducted to uncover the impact of algorithmic personalization. Algorithm audits aim to understand why certain information is shown to certain individuals or as a result of particular inputs to an algorithmic system (Sandvig et al., 2014). One of the earliest of search engine audits found no evidence of algorithmic personalization based on past search and browsing behavior on Google Search (Hannak et al., 2013). These findings are corroborated by more recent "crowdsourced audits," in which computer programs query search engines on participants' computers or accounts (Bandy, 2021; Sandvig et al., 2014). For example, Robertson et al. (2018) compared the information sources in SERPs for queries about US politics against the SERPs of participants and those of incognito browsers, concluding that algorithmic personalization of search results is relatively low and mostly dependent on query selection, time, and log-in status. Similarly, Courtois et al. (2018) found that SERPs for political search queries returned to different Google Search users do not significantly vary, indicating a lack of algorithmic personalization (cf Nechushtai et al., 2023; Robertson et al., 2023).

An alternative algorithm auditing approach employs trained virtual agents (or "sock puppets") instead of real participants to understand how specific inputs or actions lead to certain outcomes (Bandy, 2021; Sandvig et al., 2014). To our knowledge, limited research has applied this approach to the personalization of search results, especially for political topics. In the context of suicide prevention, Haim et al. (2017) examined how search engines respond to different suicide-related search queries and search histories. They show that prior suicide-harmful, suicide-preventative, or suicide-unrelated search behavior had no impact on the display of a helpline SERP feature. Regarding Google News, Haim et al. (2018) show that untrained accounts received approximately the same news articles as accounts trained according to different types of news users (also see Cozza et al., 2016; but some support for personalization found by Le et al., 2019). In general, these algorithm audits suggest that algorithmic personalization in information search is, at most, limited.

User choices

Importantly, these studies appear to underestimate the potential impact of human choices in shaping information search. Recently, Robertson et al. (2023) compared news sources shown to users by Google Search during the 2018 and 2020 US elections and the sources actively selected by these users, both overall and within Google Search. Despite being shown mostly similar SERPs, strong partisans chose to engage with more partisan and unreliable news sources, implying that user choices play a more dominant role than algorithms in the personalization of online information search.

Extensive research on selective exposure indicates that individuals tend to seek out information that aligns with their existing views when given the opportunity (Knobloch-Westerwick et al., 2015; Stroud, 2008). In political selective exposure theory, the focus has predominantly been on the selection of political content (Slechten et al., 2022). For instance, partisans exhibit preferences for different news outlets and websites (Garrett, 2009). In the realm of search engines, while the algorithmic ranking of search results drives users to predominantly interact with top-ranked results (Ulloa & Kacperski, 2023; Urman & Makhortykh, 2023), individuals also tend to opt for search results that confirm their stance on particular issues (Knobloch-Westerwick et al., 2015; Slechten et al., 2022; Westerwick et al., 2013).

Yet, the primary affordance that empowers user choices in search engines is the search query. In an initial stage before any search results can be selected, search results are determined by search queries, with Google Search specifically prioritizing "relevance" to the query in their filtering system (Google, 2023). In fact, previous studies also reported that a large part of the variation in search results about politics stems from query selection (Courtois et al., 2018; Robertson et al., 2018). Unlike the process of selecting search results, search queries allow users to choose their topics of interest rather than selecting from a set of potential choices. However, most existing algorithm auditing research assumes that individuals choose similar queries, disregarding that search query selection may vary based on individuals' preexisting political beliefs and interests. While standardized queries prove useful for isolating effects of algorithmic personalization, it is unrealistic to assume that users employ similar search strategies.

Despite its pivotal role in information search, only a handful of studies have extended the selective exposure framework to the formulation of search queries. Using survey research, van Hoof et al. (2024) found in earlier work that immigration- and climate-related search queries contain topical variations, and sometimes synonyms, which are associated with individual political characteristics, such as issue position, issue salience, and sociodemographic characteristics. How, and to what extent, these vary differs between political issues. Trielli and Diakopoulos (2022) similarly conclude that partisan searchers are interested in different topics and express these in their search queries about political candidates in the US presidential elections. Slechten et al. (2022) examined the effects of selectivity at different stages of political information search and discovered that individuals tend to construct search terms that align with their issue attitudes. Furthermore, in their study of Swiss browsing histories, Blassnig et al. (2023) find that searching for referendum-related information is uncommon and that most search terms used are rather neutral. However, they tentatively suggest user input biases in search terms about a COVID-19 referendum, acknowledging the limitations of their small sample size. Finally, Menchen-Trevino et al. (2023) similarly found that political search queries are overall neutral and do not contain overtly partisan language. Interestingly, when partisan language did occur, it was usually done by the *opposite* ideological group, which may imply a curiosity toward or verification of concepts people are unfamiliar with. While approaches and specifics vary between studies, their findings overall imply topical differences in what people with different political opinions or sociodemographic characteristics want to know about a given political topic.

Given what we know about how search engines function (Google, 2023) and conclusions from prior research (Courtois et al., 2018; Robertson et al., 2018), these varying user choices in search queries likely have a large impact on the political information presented on the SERP. Yet, this is rarely examined. To our knowledge, only Trielli and Diakopoulos (2022) analyzed the impact of distinct search queries of partisan searchers, while controlling for algorithmic personalization. The authors conclude that distinct search queries are not sufficient to lead to substantially different SERPs. That is, even with different information needs about the same political candidates, search engines will lead users to the same few information sources.

All in all, research on "filter bubbles" rarely considers that search engine users make deliberate choices that impact information exposure, most importantly *the choice for search queries*. Conversely, selective exposure research has rarely taken to investigate the extent to which such choices lead to divergent search results. In this work, we aim to combine both to answer the following research question:

> RQ1: To what extent do user choices in search queries and algorithmic personalization lead to divergent information sources in political Google Search results?

Dimensions of divergent search results

The extent to which search engine results diverge as a result from algorithmic personalization or user choices in search queries can be evaluated along several dimensions, including the (type of) information sources and features present on the SERP, that provide additional insight into how search results may diverge. Previous research on individual variation in search engine results have mostly focused on search engines' selection and ranking of information sources. Some have understood this as the (ranked) overlap of exact URLs or information sources (Haim et al., 2017; Hannak et al., 2013; Robertson et al., 2018; Trielli & Diakopoulos, 2022). Understanding whether search engines drive different individuals to different information sources matters because not all sources are made equal: They may vary in terms of political bias, credibility, objectivity, expertise, and other characteristics.

Scholarly attention has mostly been paid to the categorization of information sources along partisan or ideological lines (Hu et al., 2019; Robertson et al., 2023; Robertson et al., 2018), particularly in the United States. While this approach is relevant in the United States and similar media systems characterized by a high degree of polarization, media outlets in the Netherlands and many other Western European countries do not align distinctly along ideological lines.

To delve further into the nuances of personalization in search results, it is essential to consider these different *types of* information sources. Political queries yield results that encompass mainstream news sources, civil society organizations, governmental sources, Wikipedia, political party websites, and social media (Courtois et al., 2018; Puschmann, 2019; Steiner et al., 2022; Trielli & Diakopoulos, 2019; Unkel & Haim, 2021), depending on the query and search engine. Notably, the informational value of different types of sources may vary substantially. For instance, quality news sources are expected to adhere to journalistic ideals of objectivity and neutrality when reporting on divisive political issues, unlike alternative news sources like social media. Background information sources such as Wikipedia offer neutral information that is not overtly one-sided or opinionated, while governmental sources predominantly provide information on policy and governmental services. Consequently, if individuals are directed to different types of information sources based on their choice of search queries or previous interactions with the search engine, the composition of their information diets may differ substantially when seeking information about the *same* political issues.

Furthermore, we extend existing research by considering the impact of user choices and algorithmic personalization on the display of SERP features (e.g., Featured Snippet, Knowledge Panel, Top Stories, see Supplementary Figure SM1 for an illustration). Despite being standard components of modern SERPs (Oliveira & Lopes, 2023), these have mostly been neglected by existing research (see also Robertson et al., 2018). These features present information in boxes separated from regular search results and often take top-ranked positions at the expense of regular search results. Coupled with individuals' selective interaction with the top search results (Pan et al., 2007; Urman & Makhortykh, 2023), these features draw attention away from other search results (Epstein et al., 2022; Gleason et al., 2023). Importantly, these SERP features prioritize different types of information. For instance, Top Stories highlights news, particularly from mainstream or legacy media (Trielli & Diakopoulos, 2019), and the Twitter feature displays recent tweets related to specific topics or by notable users. Google Search's newest AI Overviews feature provides comprehensive AI-generated answers to queries directly on the SERP.

The display of SERP features may be influenced by different user choices or algorithmic personalization, resulting in differential information about the same political topic. Tentative evidence suggests that the display of these features is related to the search query, as seen in studies on suicide prevention features (e.g., Arendt et al., 2020; Haim et al., 2017; Haim et al., 2021). Robertson et al. (2018) investigated SERPs returned to political searches in the United States and found that substantial variation in the presence of SERP features by search query, but it was not significantly affected by algorithmic personalization. However, the extent to which certain individuals are shown specific SERP features, depending on their search queries or previous interactions with the platform, remains largely unknown.

We ask the following research question:

> RQ2: How do user choices in search queries and algorithmic personalization manifest in distinct information sources, types of information sources, and SERP features in political Google Search results?

Data and methods

ABT (Haim, 2020) is an algorithm auditing technique in which "researchers collect data by creating computer programs which impersonate users who then test the algorithm" ("sock puppet audit," Bandy 2021, p. 74; see also Sandvig et al., 2014). Although more artificial than observational studies, ABT allows for isolating specific inputs to an algorithmic system that are difficult to control in other settings while improving reliability and ecological validity (Bandy, 2021; Haim, 2020). In our experiment, virtual agents simulate users by conducting Google searches, and we compare resulting SERPs.

The collections of the SERPs was approved by the Ethical Review Board of the University of Amsterdam. Data and materials for this article are available on OSF,[2] and the data preparation and analysis scripts are on GitHub.[3] The search query data used by van Hoof et al. (2024) are publicly available on OSF[4] (Araujo et al., 2020).

Research design

To operationalize algorithmic personalization, we consider the *past* searches conducted by an agent. Google Search uses search history to curate the SERP for future searches (Google, 2023). Therefore, past search queries are expected to signal specific interests to the search engine algorithm. User choice is operationalized as the *current* search query.

We conduct two separate experiments on immigration and climate change with the following experimental designs. Immigration: 3 (User choice: Pro, Neutral, Anti) × 6 (Algorithmic personalization: Pro, Neutral, Anti, Mixed, Unrelated, None) and climate change: 3 (User choice: High, Neutral, Low) × 6 (Algorithmic personalization: High, Neutral, Low, Mixed, Unrelated, None). Each condition encompasses a set of search queries.

We construct the search terms based on van Hoof et al. (2024), who identified clusters of Dutch individuals using distinct immigration and climate-related search queries collected via a survey, which were associated with issue attitudes, issue importance, and sociodemographic characteristics. For immigration, we distinguish between *Pro* and *Anti* search behavior, reflecting the most frequent search terms used by those with opposing attitudes toward immigration. For climate, we distinguish search behavior based on the perceived issue importance: *Low issue importance* and *High issue importance* (for brevity reasons, referred to as *Low* and *High*), reflecting the search terms most frequently used by those groups. (Consult Supplementary Table SM1 for details on how these conditions align with the findings presented in van Hoof et al. (2024)). Extending van Hoof et al. (2024), the *Neutral* condition consists of queries that do not reflect any stance and were used by multiple groups. These conditions apply to both algorithmic personalization and user choice.

The search queries for these conditions were selected based on van Hoof et al. (2024)'s study findings and data following a systematic procedure to ensure reliability, which is further detailed in the Supplementary material. For each condition, three search queries are reserved to measure user choice (indicated by *anti_1, …, anti_3, … pro_1, …, pro_3*), while the remainder is used for algorithmic personalization (see Table 1 for an overview).

Additionally, three other algorithmic personalization conditions were created. *Mixed* search behavior combines queries from both Pro and Anti or High and Low conditions equally, capturing seeking divergent opinions (Menchen-Trevino et al., 2023). *Unrelated* search behavior indicates an interest in (news) topics other than immigration and climate. It is constructed by extracting the top 20 most frequent named entities in Dutch news headlines.[5] Finally, the *None* condition represents no search history, measuring the absence of algorithmic personalization.

Table 1. Overview of experimental conditions and search terms

Condition	Algorithmic personalization	User choice
Immigration		
Pro	vluchteling vluchtelingen nederland vluchtelingen in nederland … ($n = 14$)	vluchtelingen (*pro_1*) vluchtelingencrisis (*pro_2*) vluchtelingenproblematiek (*pro_3*)
Neutral	immigranten immigratie nederland migratie … ($n = 13$)	immigratie (*neutral_1*) imigranten (*neutral_2*) immigratiecijfers (*neutral_3*)
Anti	opvang asielzoekers illegalen asiel beleid … ($n = 10$)	asielzoekers (*anti_1*) azc (*anti_2*) criminaliteit onder asielzoekers (*anti_3*)
Mixed	vluchteling vluchtelingen nederland opvang asielzoekers … ($n = 20$)	–
Climate		
High	zonnepanelen politiek klimaat klimaattop … ($n = 27$)	klimaat politiek (*high_1*) politiek en klimaat (*high_2*) klimaatakkoord (*high_3*)
Neutral	klimaatverandering klimaat verandering klimaatveranderingen … ($n = 4$)	klimaat (*neutral_1*) milieu (*neutral_2*) klimaat veranderingen (*neutral_3*)
Low	opwarming opwarming van de aarde stijging zeespiegel … ($n = 26$)	opwarming aarde (*low_1*) zeespiegel stijging (*low_2*) aarde opwarming (*low_3*)
Mixed	opwarming opwarming van de aarde zonnepanelen … ($n = 52$)	–
Both issues		
Unrelated	npo nederland rtl … ($n = 25$)	–
None	–	–

Note. The full list of search terms used for each algorithmic personalization condition are included in the Supplementary Tables SM2 (immigration) and SM3 (climate).

Data collection

We used ScrapeBot (Haim 2020; see also the Github[6]), an open-source Selenium-based tool, to employ ABT. ScrapeBot, which allows for the simulation of human interaction with web pages in a visual browser, has been utilized in similar research designs (Haim et al., 2017; Haim et al., 2018).

We created 114 agents for each of the 18 conditions per political issue ($N_{immigration} = 2052$; $N_{climate} = 2052$). As one server can only conduct a single search for one agent at a time, we parallelize the data collection across 19 servers (i.e., 108 agents per server), ensuring each condition is equally represented on every server. This approach allows us to scale up the number of agents while ensuring limited time-related influence on the search results.

The data collection is performed consecutively for each of the two issues. The data collection procedure is divided into two phases: *training* and *testing* (see Figure 1 for an overview of the data collection procedure). During the *training phase*, we created a search history for each agent by conducting multiple Google searches with cookie collection turned on. The program randomly selects an agent, retrieves their saved cookies, and randomly selects a search query from the agents' assigned set of search queries. It then searches this query on www.google.nl, retrieves autocomplete predictions, the first SERP's HTML, and the cookies. It then moves on to the next agent to repeat the process. For an agent's first run, a fresh browser environment (without histories or cookies) is initiated, the cookie statement is accepted, and the training procedure is followed, after which the agents' cookies are stored in a central database, and retrieved and added to the browser environment for every next run that agent performs after that.

Google identifies an agent via its associated cookies rather than creating individual Google accounts for each agent, following methodology for similar studies (Haim et al., 2017; Haroon et al., 2023). Google Search's documentation suggests that while Google Search personalizes search results based on activity saved in Google accounts (Google, 2023), it also personalizes search results based on cookies, even for users who are not logged in but have accepted the cookie statement (Google, 2024). The training phases were ended after 13 days (Immigration: August 18–31, 2023) or 11 days (Climate: September 4–15, 2023), during which each agent was trained $M_{immigration} = 257.8$ and $M_{climate} = 217.6$ times.[7]

The *testing phase* began within an hour the training phase ending for *all* agents. Agents trained according to the five different search histories are assigned to user choice conditions. During the testing phase, each agent conducts three testing searches, one for each of the three search queries. For the condition without search history, we use browser environments without any cookies loaded. These agents lack any search history, preventing personalization based on past searches.

During the testing procedure, the program randomly selects an agent and query combination, loads their training phase cookies (if applicable), navigates to www.google.nl, performs a search using a query from the agent's assigned set, retrieves the autocomplete predictions and the first SERP's HTML, and proceeds to the next agent. The testing phases ran on one day (Immigration: August 31, 2023; Climate: September 15, 2023).

Following best practices for ABT (Schwabl et al., 2024) and based on insights from Google Search's documentation (Google, 2023), previous studies (Haim, 2020), and pretesting, we control for location, language settings, and browser settings in all searches: We set browser-language settings to Dutch, use servers located in Amsterdam, and use the same display size, browser type (Chrome), and use the most common Dutch user-agent string. While it is difficult to fully control for time-related influence on the search results, we minimized the time difference of testing searches between agents by parallelization across 19 servers: All testing searches are completed within 3 hours and 39 minutes.

Data preparation

Using the WebSearcher parsers[8] (Robertson & Wilson, 2020), we extracted the URLs, textual content (e.g., title,

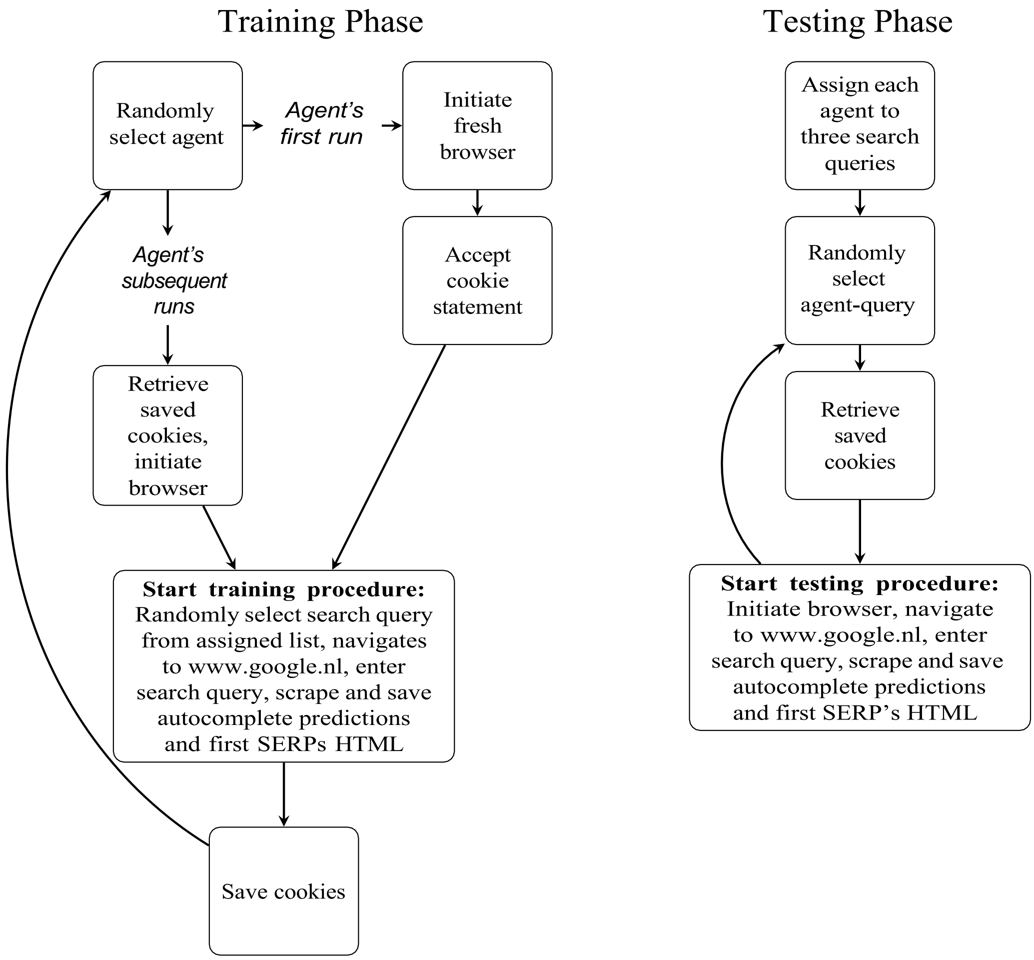

Figure 1. Overview of data collection procedure.

Note. The data collection procedure is parallelized across 19 servers.

snippet text), and SERP features, and their (sub)rank positions from the HTML of each SERP collected during the testing phase. This process created a dataset where each SERP is depicted by multiple rows, each corresponding to a search result. We identified regular search results and 12 SERP features. On average, each SERP contained 23.19 search results ($SD = 4.89$), 12.18 components ($SD = 1.19$) and 4.11 distinct SERP features ($SD = 1.61$).[9] Less than 1% of cases involve unparseable SERP features, and in less than 1% of both climate and immigration cases content of features could not be parsed.

We extracted the information source (domain) from all URLs. The *type of information source* was determined using an existing list of domains categorizations (Loecherbach, 2023). Initially, this categorized approximately 80% of search results. An additional coding round was conducted to achieve full coverage. The sources are classified into news, background information (about public events and figures, e.g. Wikipedia, governmental or non-profit organizations), gateway (e.g., social media, search engines), and other websites unrelated to news and politics.

Dissimilarity metrics

To study whether user choices in search queries and algorithmic personalization lead to divergent search results, we need metrics to assess the (dis)similarity of search results across experimental conditions. Pairwise similarity metrics, commonly used in informational retrieval literature, are suitable for this purpose. An appropriate similarity measure should account for two curation decisions by search engines: the *selection* of search results from the index, and their *ranking*. Two SERPs can be identical in selection, but differently ranked. This is crucial, given the strong order effects in individuals' interaction with search results, with lower-ranked results being less likely to be selected (Urman & Makhortykh, 2023).

We use Ranked-Biased Overlap (RBO), which is developed for comparing search results (Webber et al., 2010) and is increasingly employed in similar research (Makhortykh et al., 2020; Robertson et al., 2018; Urman et al., 2022). RBO addresses both selection and ranking by assigning greater weight to top results. The p (persistence) parameter determines this weighting, with smaller values emphasizing top results more. Following prior research (Urman et al., 2022), we computed RBO ($p = .8$) for a ranking-sensitive metric. To measure *differences* in SERPs, we inverted the RBO (1 − RBO) to obtain Inverted RBO (IRBO), ranging from 0 (identical) to 1 (completely dissimilar). IRBO is computed on the sources listed on the SERP, excluding search results without information sources like People Also Ask and images, across all possible SERP pairs, resulting in approximately 19 million comparisons per issue.

As a robustness check, we replicate our analyses using IRBO ($p = .95$), which is less attuned to ranking (Urman et al., 2022), and (Inverted) Jaccard Index, another common similarity measure for search results (e.g., Hannak et al., 2013; Kliman-Silver et al., 2015; Makhortykh et al., 2020; Puschmann, 2019). These additional analyses are detailed in the Supplementary material.

Results

Algorithmic personalization versus user choices in search queries

Figures 2 (immigration) and 3 (climate change) illustrate the average dissimilarity of information sources on SERPs returned to different search queries, grouped by user choice conditions on both axes. To assess whether algorithmic personalization resulted in divergent search results, we can examine the standard deviations of the mean dissimilarity scores across algorithmic personalization conditions within each search query-pair. These are presented in brackets in Figures 2 and 3 (see Supplementary Figures SM2 (immigration) and SM3 (climate change) for dissimilarity scores by algorithmic personalization-pairs per search query). For both issues and across all search queries, the standard deviations consistently approach zero, indicating minimal variation in dissimilarity scores across different algorithmic personalization conditions. This implies that the differences in the selection and ranking of information sources on the SERP can be attributed very little, if at all, to the agents' differently trained search histories. Hence, algorithmic personalization did not lead to divergent information sources.

To assess the impact of user choices in search queries, we compare the dissimilarity scores between *different* search queries (i.e., the scores below the diagonal) to the scores within each search query (i.e., the scores on the diagonal). We also examine the scores of search queries within the same or different user choice conditions (i.e., clustered together on the axes).

For immigration-related queries (Figure 2), we observe a relatively high dissimilarity between search results for different search queries. The scores between different immigration-related search queries range from .75 to .99, indicating that 75% to nearly all of the information sources near the top of the SERPs are unique to each query. This level of dissimilarity is substantially higher than the dissimilarity observed within the same search query, which ranges from .09 to .56. Some dissimilarity in search results for the same query is likely, possibly due to factors like randomization by Google Search (Urman et al., 2022), minor time differences in data collection, or other unknown factors. The dissimilarity scores belonging to the same user choice conditions are *not* substantially lower than between queries, indicating that search results are unique at the granular level of search queries rather than aggregated user choices.

These patterns are less clear for climate-related search results (Figure 3). In line with the immigration search results, we generally observe higher dissimilarity scores between different search queries compared to within the same search queries. However, there are a few exceptions. The dissimilarity between search results returned to *low_1*, *low_3*, and *neutral_3* is relatively low, equivalent to the level of dissimilarity within the same search queries. This indicates that these search result pages share many of the same information sources in top-ranked positions. Similar to immigration-related search results, these scores do not clearly cluster at the level of user choices. While the similarity between search results returned for *low_1* and *low_3* might suggest clustering at the user choice level, this is rather explained by the nearly identical nature of these search queries (i.e., same words in different order).

The Supplementary material presents the analyses using two alternative metrics: Inverted Jaccard Index and IRBO ($p = .95$). The patterns of dissimilarity are similar between IRBO ($p = .80$), presented in Figures 2 and 3, and Inverted Jaccard Index, while for IRBO ($p = .95$) the dissimilarity between search results for the *same* queries is substantially higher. This indicates a high degree of overlap between search results returned to the same search query, especially among top-ranked results, and that differences in search results for identical queries occur in the "long tail." In other words, while the most relevant and highly ranked search results are similar, the results start to diverge with less relevant and lower-ranked search results, which is consistent with findings from previous research (Steiner et al., 2022; Urman et al., 2022).

SERP composition

The previous results described *the extent to which* search results vary due to user choices in search queries and algorithmic personalization (RQ1); this section examines *how* the SERPs varies with different search queries, specifically focusing on the information source, type of source, and the display of SERP features (RQ2). Since algorithmic personalization showed no impact on search results, we focus on user choices in search queries specifically by selecting only agents without search history. In this section, we present a weighted proportion to account for rank order, assigning weights to search results inversely proportional to their rank, giving higher weights to top-ranked search results (i.e., $\frac{1}{(rank+1)}$), and summing these weights for each source (type).

The large variation in source types and SERP features displayed in Figure 4 reflects the high dissimilarity scores between immigration queries. A few points stand out. While news sources are prevalent for two search queries (i.e., *pro_1* and *anti_1)* due to the display of the Top Stories feature, news is mostly absent or have much a lower weighted proportions for other queries. Furthermore, the SERPs for *anti_2*, *anti_3*, and *neutral_1* contain a relatively high weighted proportion of gateway websites like social media. Specifically, Twitter emerges as a frequent top-ranked source for *anti_2* and *neutral_1* (see Table 2), related to the presence of a Twitter feature showcasing "top tweets" about the queried topic. A manual inspection of these tweets reveals that those for *anti_2* contain mostly news headlines, whereas the tweets for *neutral_1* convey strong negative opinions about immigration. Furthermore, *anti_3* features a scholarly articles SERP feature, making google.nl a frequent and top-ranked gateway source.

The SERP composition for climate queries is presented in Figure 5, and it shows that the search results generally consist of mostly background information sources (e.g., governmental websites, Wikipedia, see Table 3) or websites unrelated to news or politics. Unlike immigration queries, gateway websites and, to a lesser extent, news, are rarely frequent and highly ranked parts of the search results. An exception are

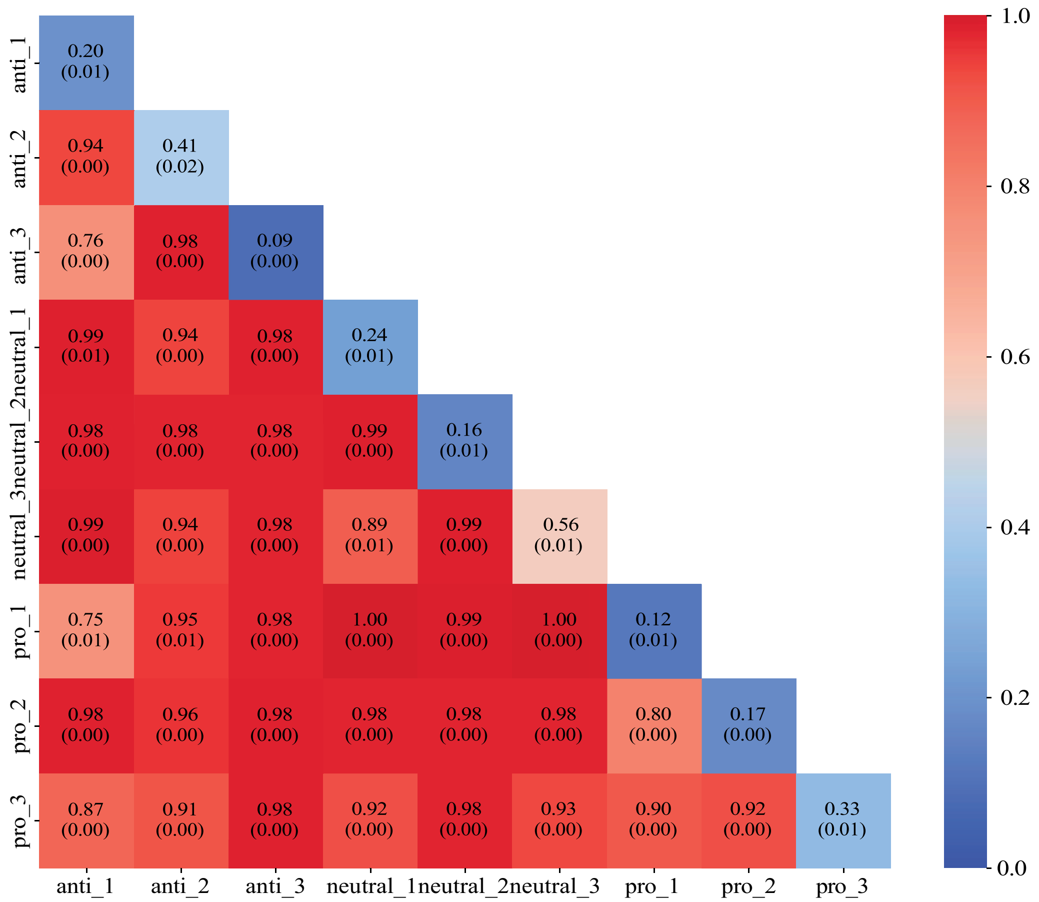

Figure 2. Immigration: Average dissimilarity of sources between and within search queries.

Note. Values represent the average IRBO ($p = 0.8$) for each search query-pair, grouped by user choice condition. The scores on the diagonal represent the dissimilarity within search queries, while the scores below the diagonal represent the similarity between search queries. Values in brackets represent the standard deviation of the mean dissimilarity scores for algorithmic personalization conditions, grouped by search query-pair. The standard deviations are close to zero, indicating little impact of algorithmic personalization.

the search results for *neutral_1*, which shows a large share of news sources due to the Top Stories feature.

Discussion and conclusion

Search engines are a major pathway to political information and news (Wojcieszak et al., 2022). Given their significant influence on political information exposure and behavior (Epstein & Robertson, 2015; Pan et al., 2007), this study addressed concerns about the extent to which user choices and algorithmic personalization lead to divergent search results about political issues (RQ1).

Our findings confirm prior research on filter bubbles in information search (e.g., Courtois et al., 2018; Puschmann, 2019): Algorithmic personalization alone does not lead to divergent information sources in political search results. Put simply, identical queries produce the same search results, regardless of variations in users' search histories. Instead, our study shows that the search queries people formulate play a key role in shaping their exposure to search results. When (artificial) users input different queries about the same political issue, they encounter largely distinct information sources. This pattern is evident in searches related to both immigration and climate, although search results for climate-related queries tend to be more similar.

These findings have several important implications. First, these findings are important in the ongoing discourse around information exposure via search engines and other digital platforms, which often emphasizes the influence of technology. Previous algorithm auditing studies typically rely on a set of standardized queries generated by researchers to examine political search results. While these studies are valuable for understanding general patterns in search engine outputs and the influence of algorithmic factors, they might not accurately reflect the variety of search queries used in real life. Our study adopted a user-centered approach by basing the research design on a typology of political queries derived from real users, which can better address questions related to how people actually seek out and encounter political information through search engines.

Second, previous research indicates that the use of different search queries is associated with individuals' political attitudes (e.g., Slechten et al., 2022; Trielli & Diakopoulos, 2022; van Hoof et al., 2024), underscoring the importance of considering a diversity of search terms. However, our study did not find support for the clustering related to political attitudes, as identified in previous research (van Hoof et al., 2024), influencing the search results in the same way. Instead, we discovered that the formulation of search queries is the primary factor driving the diversity of search results

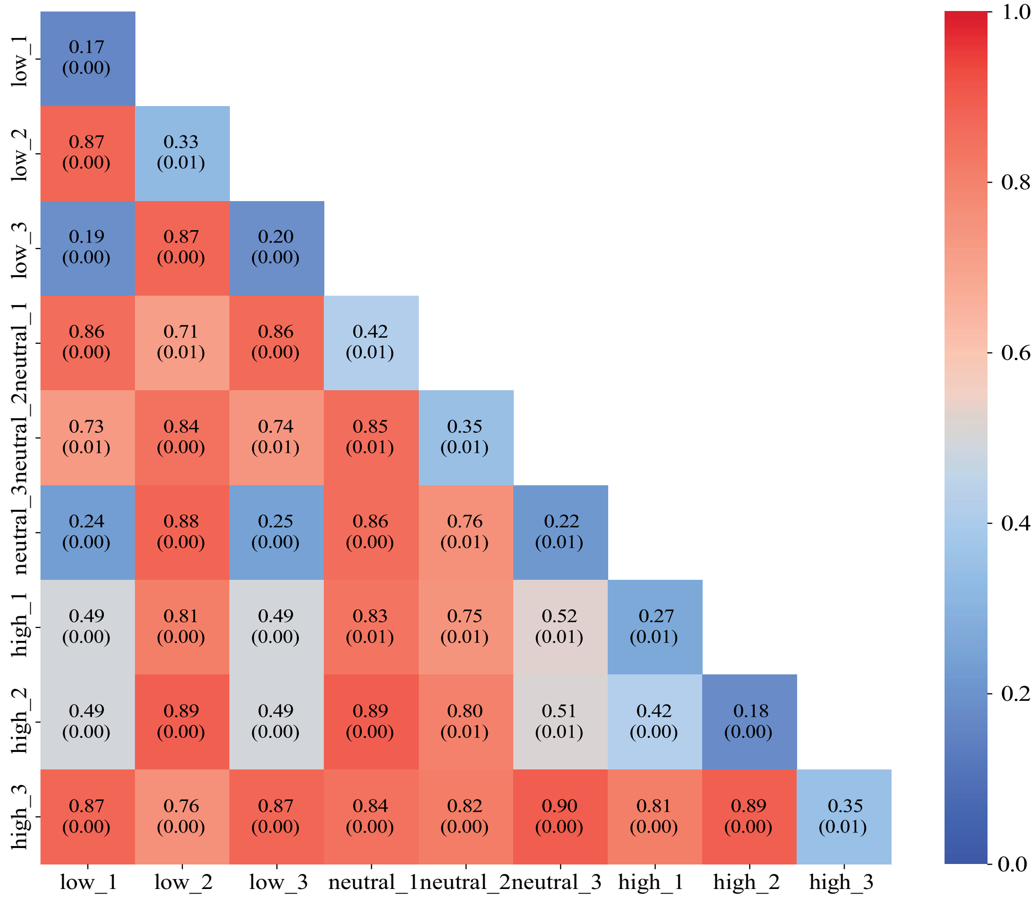

Figure 3. Climate: Average dissimilarity of sources between and within search queries.

Note. Values represent the average IRBO ($p = 0.8$) for each search query-pair, grouped by user choice condition. The scores on the diagonal represent the dissimilarity within search queries, while the scores below the diagonal represent the similarity between search queries. Values in brackets represent the standard deviation of the mean dissimilarity scores for algorithmic personalization conditions, grouped by search query pair. The standard deviations are close to zero, indicating little impact of algorithmic personalization.

encountered by users. This might suggest that for these two issues, clusters of people are not pronounced enough when it comes to search queries, ultimately not leading to distinct clusters in search results. Alternatively, this could be an artifact of the design used by van Hoof et al. (2024) or how we adapted their design to our research.

Regardless, these intriguing findings call for replication and extension. This study is a case study of Google Search on searches for two political topics during a non-campaign period in the Netherlands. The specific search queries and search results may vary based on factors like timing, country, topic, and search engine. For instance, our findings differ from Trielli and Diakopoulos (2022) who found a mostly similar selection of search results for divergent partisan search queries about US political candidates, possibly because dominant news stories and outlets influence results the day before elections. This indicates that further research is necessary to understand how these mechanisms unfold in different contexts.

Third, our findings complement existing scholarship on user selections of search results identifying ranking (Pan et al., 2007; Ulloa & Kacperski, 2023; Urman & Makhortykh, 2023) and confirmation biases (i.e., political selective exposure)(Knobloch-Westerwick et al., 2015; Westerwick et al., 2013). We contribute to this literature by focusing on an earlier stage in the search process: Before users can select any search result, a query needs to be formulated that impacts the results that are even available for selection. While recent studies have given more attention to political search queries (e.g., Blassnig et al., 2023; Menchen-Trevino et al., 2023; Slechten et al., 2022; van Hoof et al., 2024), they had rarely extended their research to how they translate into information exposure on the search result page. In future research, researchers could combine multiple steps in the search process, such as follow whether user selections of search results further amplify biases introduced by query formulations (also see Slechten et al., 2022).

In the second part of our study, we demonstrated that search queries and search topics (immigration or climate) can yield different distributions of information sources and SERP features (RQ2). This is important because different types of sources, such as news or social media content, and the SERP features that highlight them, may inform users in varied ways. Furthermore, (AI-generated) SERP features have become integral parts of search engines, often making the SERP more influential than the linked pages (Epstein et al., 2022; Gleason et al., 2023). Interestingly, these features tend to be binary: They either consistently appear for specific keywords or topics, or not at all. This pattern highlights the need to better understand how users contribute to these patterns.

These findings offer qualitative insights into how varying search queries on similar topics influence the type of sources

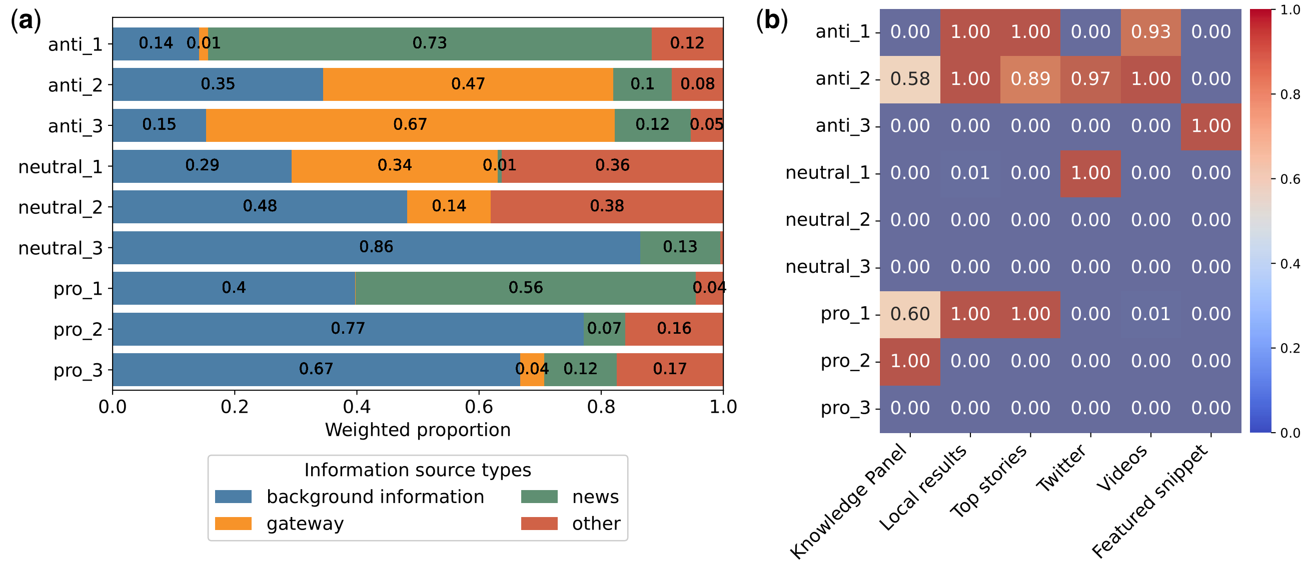

Figure 4. Immigration: SERP composition. (a) Information source types. (b) SERP features.

Note. (a) Weights are assigned to search results inversely proportional to their rank. (b) Values indicate the share of SERPs containing a specific SERP feature for each search query.

Table 2. Immigration: Most frequent information sources per user choice condition by weighted proportion

anti_1	prop^w	type	*anti_2*	prop^w	type	*anti_3*	prop^w	type
ad.nl	0.13	news	twitter.com	0.41	gateway	google.nl	0.63	gateway
nrc.nl	0.11	news	coa.nl	0.20	bg info	rijksoverheid.nl	0.10	bg info
rtlnieuws.nl	0.11	news	wikipedia.org	0.12	bg info	quest.nl	0.05	other
volkskrant.nl	0.11	news	youtube.com	0.06	gateway	vpro.nl	0.04	news
looopings.nl	0.10	other	azczutphen.nl	0.04	other	nos.nl	0.03	news
neutral_1	**prop^w**	**type**	***neutral _2***	**prop^w**	**type**	***neutral _3***	**prop^w**	**type**
twitter.com	0.34	gateway	cbs.nl	0.40	bg info	cbs.nl	0.54	bg info
ind.nl	0.31	other	encyclo.nl	0.13	gateway	rijksoverheid.nl	0.13	bg info
rijksoverheid.nl	0.10	bg info	woorden.org	0.09	other	nos.nl	0.08	news
cbs.nl	0.08	bg info	rijksoverheid.nl	0.08	bg info	adviesraadmigratie.nl	0.07	bg info
wikipedia.org	0.07	bg info	uu.nl	0.07	other	europa.eu	0.06	bg info
pro_1	**prop^w**	**type**	***pro_2***	**prop^w**	**type**	***pro_3***	**prop^w**	**type**
ad.nl	0.14	news	wikipedia.org	0.48	bg info	rijksoverheid.nl	0.40	bg info
telegraaf.nl	0.14	news	vluchteling.nl	0.14	bg info	universiteitleiden.nl	0.13	other
trouw.nl	0.14	news	rodekruis.nl	0.08	other	amnesty.nl	0.10	bg info
nu.nl	0.13	news	universiteitleiden.nl	0.06	other	unhcr.org	0.08	bg info
wikipedia.org	0.12	bg info	nrc.nl	0.05	news	bnnvara.nl	0.07	news

Note. prop^w: Weighted proportion. Weights are assigned to search results inversely proportional to their rank, giving higher weights to higher ranked search results (i.e., $\frac{1}{rank+1}$), which are then summed for each source.

users encounter when seeking political information. While interesting, these results also call for replications to see how robust these findings are, such as across contexts. Moreover, with our focus on source (types), we could not determine the extent to which the SERPs actually provided different information. To uncover these patterns, future studies could develop new metrics to analyze the dissimilarity of textual content presented on SERPs or the underlying web pages.

This study has limitations in scope and method. Our methodological approach uses survey-based search queries, which is an important step toward understanding how search query choice may affect political information exposure. However, survey-based search queries may not fully capture real-life search behavior (Blassnig et al., 2023), which affects the study's ecological validity.

Furthermore, our implementation of ABT enabled us to isolate the effects of search history on search results, gaining experimental control over other more ecological valid alternatives like crowdsourced algorithm audits. This comes with several implications. First, we focused on one signal of algorithmic personalization: search history. Despite efforts to enhance ecological validity, including using real queries and an

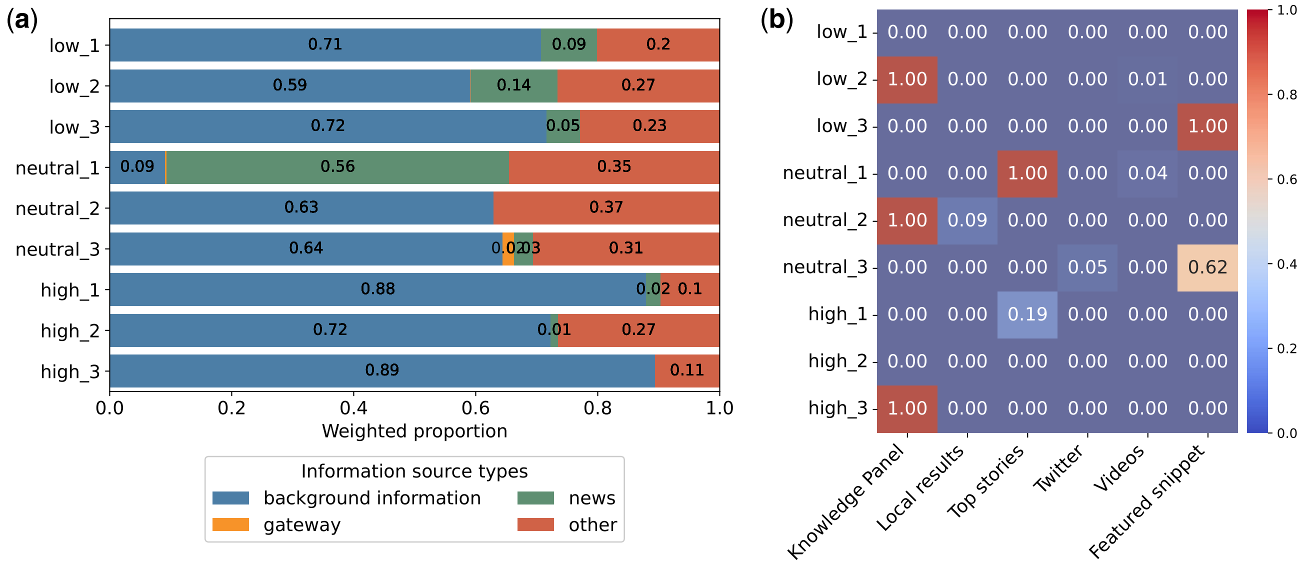

Figure 5. Climate: SERP composition. (a) Information source types. (b) SERP features.

Note. (a) Weights are assigned to search results inversely proportional to their rank. (b) Values indicate the share of SERPs containing a specific SERP feature for each search query.

Table 3. Climate: Most frequent information sources per user choice condition by weighted proportion

low_1	prop^w	type	*low_2*	prop^w	type	*low_3*	prop^w	type
klimaatakkoord.nl	0.37	bg info	wikipedia.org	0.36	bg info	klimaatakkoord.nl	0.25	bg info
klimaat.be	0.11	bg info	knmi.nl	0.23	other	europa.eu	0.17	bg info
wwf.nl	0.10	other	rijksoverheid.nl	0.08	bg info	klimaat.be	0.12	bg info
nos.nl	0.09	news	nos.nl	0.08	news	wwf.nl	0.10	other
wikipedia.org	0.08	bg info	deltaprogramma.nl	0.06	bg info	rijksoverheid.nl	0.07	bg info
neutral_1	**prop^w**	**type**	***neutral_2***	**prop^w**	**type**	***neutral_3***	**prop^w**	**type**
telegraaf.nl	0.22	news	wikipedia.org	0.44	bg info	rijksoverheid.nl	0.43	bg info
knmi.nl	0.17	other	milieucentraal.nl	0.12	other	wwf.nl	0.13	other
ad.nl	0.14	news	rijksoverheid.nl	0.06	bg info	klimaatadaptatienederland.nl	0.12	bg info
fd.nl	0.12	news	wiktionary.org	0.05	other	europa.eu	0.09	bg info
warmte365.nl	0.04	other	apple.com	0.04	other	wur.nl	0.06	other
high_1	**prop^w**	**type**	***high_2***	**prop^w**	**type**	***high_3***	**prop^w**	**type**
rijksoverheid.nl	0.61	bg info	rijksoverheid.nl	0.52	bg info	wikipedia.org	0.35	bg info
wwf.nl	0.08	other	wwf.nl	0.10	other	klimaatakkoord.nl	0.23	bg info
groenlinks.nl	0.07	bg info	klimaatlabelpolitiek.nl	0.07	bg info	rijksoverheid.nl	0.20	bg info
d66.nl	0.06	bg info	d66.nl	0.06	bg info	vng.nl	0.05	bg info
klimaatlabelpolitiek.nl	0.06	bg info	ipsos.com	0.05	other	emissieautoriteit.nl	0.05	other

Note. prop^w: Weighted proportion. Weights are assigned to search results inversely proportional to their rank, giving higher weights to higher ranked search results (i.e., $\frac{1}{rank+1}$), which are then summed for each source.

extensive training phase, it remains uncertain whether this was sufficient for search engine algorithms to pick up on. Second, as we used non-logged in virtual agents in our implementation, our measurement of algorithmic personalization is restricted to that which occurs via cookies, and we cannot account for personalization based on activity saved in Google accounts. Third, while our interest was in understanding the "long-term" impact of search history on information exposure, as suggested by the filter bubble hypothesis, the short-term or "carry-over effects" of search history (Hannak et al., 2013), that is, impact of searches within a brief time window, warrant further research. Therefore, we encourage future research to explore other and related signals of algorithmic personalization in information search. Fourth, while it is impossible to control for all potentially confounding factors that impact search results, we followed best practices for ABT and controlled for known technical confounding variables (Schwabl et al., 2024). Yet, there are potentially other unknown factors that may have influenced how realistic our virtual agents were, such as multiple searches coming from the specific IP address of the servers we used. Regardless, data access facilitated by platforms like Google Search, and methodological innovations when such research opportunities remain unavailable, are required to further understand the interplay between users, algorithms, and search results in future studies.

While we should stay conscious of how search engines tailor information based on our past online behavior, especially for critical topics like politics, our findings deflate the notion of algorithm-driven information bubbles in Google Search. Our study emphasizes that focusing solely on algorithms may lead to an underestimation of user-driven effects. To understand how search engines deliver political information, it is essential to account for human choices in search queries. We suggest that future research in this field prioritizes human choices in information search rather than controlling for them.

Supplementary material

Supplementary material is available at *Journal of Computer-Mediated Communication* online.

Data availability

The data used for this article are available in the Open Science Framework at https://osf.io/h3a94/.

Funding

The data collection of search queries was supported by the Amsterdam School of Communication Research and its Digital Communication Methods Lab. The experiment was carried out on the Dutch national e-infrastructure with the support of the SURF Foundation (Grant No. EINF-5514).

Conflicts of interest: The authors declare there is no conflict of interest.

Open science framework badges

Open Materials

The components of the research methodology needed to reproduce the reported procedure and analysis are publicly available for this article.

Open Data

Digitally shareable data necessary to reproduce the reported results are publicly available for this article.

Notes

1. https://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html
2. https://osf.io/h3a94/
3. https://github.com/mariekevh/ItMattersHowYouGoogleIt
4. https://osf.io/yu64r/
5. News headlines (n = 296, 355) are obtained via the AmCAT (Amsterdam Content Analysis Tool, https://github.com/amcat) published between April 1st, 2021 and June 30, 2022 by the six major Dutch newspapers in the Netherlands (De Telegraaf, NRC Handelsblad, De Volkskrant, Algemeen Dagblad, Trouw, and Het Financieele Dagblad).
6. https://github.com/MarHai/ScrapeBot
7. Due to some initial technical errors, there is a minor two-day difference between the training phases of both issues. This discrepancy is unlikely to impact our results, as algorithmic personalization has minimal influence (see Results section of this chapter). Additionally, in extremely rare instances, training phase runs failed (24 and 38 cases for immigration and climate, respectively). Since these occurred around the same time and were not isolated to specific server or agent, we suspect platform-related issues (e.g., extended loading times). We also do not expect these unsuccessful runs to affect our findings.
8. https://github.com/gitronald/WebSearcher
9. Search results are different from components in that a search result, e.g. a news item, can be included within a higher-level component, e.g. Top Stories.

References

Araujo, T., Boukes, M., Trilling, D., van Hoof, M., Wald, R., & Zhang, D. (2020). Communication in the Digital Society Survey in the Netherlands. https://doi.org/10.17605/OSF.IO/YU64R

Arendt, F., Haim, M., & Scherr, S. (2020). Investigating Google's suicide-prevention efforts in celebrity suicides using agent-based testing: A cross-national study in four European countries. *Social Science & Medicine*, *262*, 112692. https://doi.org/10.1016/j.socscimed.2019.112692

Bandy, J. (2021). Problematic machine behavior: A systematic literature review of algorithm audits. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–34. https://doi.org/10.1145/3449148

Blassnig, S., Mitova, E., Pfiffner, N., & Reiss, M. V. (2023). Googling referendum campaigns: Analyzing online search patterns regarding Swiss direct-democratic votes. *Media and Communication*, *11*(1), 19–30. https://doi.org/10.17645/mac.v11i1.6030

Boomgaarden, H. G., & Vliegenthart, R. (2007). Explaining the rise of anti-immigrant parties: The role of news media content. *Electoral Studies*, *26*(2), 404–417. https://doi.org/10.1016/j.electstud.2006.10.018

Cardenal, A. S., Aguilar-Paredes, C., Galais, C., & Pérez-Montoro, M. (2019). Digital Technologies and Selective Exposure: How Choice and Filter Bubbles Shape News Media Exposure. *The International Journal of Press/Politics*, *24*(4), 465–486. https://doi.org/10.1177/1940161219862988

Courtois, C., Slechten, L., & Coenen, L. (2018). Challenging Google Search filter bubbles in social and political information: Disconforming evidence from a digital methods case study. *Telematics and Informatics*, *35*(7), 2006–2015. https://doi.org/10.1016/j.tele.2018.07.004

Cozza, V., Hoang, V. T., Petrocchi, M., & Spognardi, A. (2016). Experimental Measures of News Personalization in Google News. In S. Casteleyn, P. Dolog, & C. Pautasso (Eds.), *Current Trends in Web Engineering* (pp. 93–104). Springer International Publishing. https://doi.org/10.1007/978-3-319-46963-8_8

Diakopoulos, N., Trielli, D., Stark, J., & Mussenden, S. (2018). I vote for—How search informs our choice of candidate. In M. Moore & D. Tambini (Eds.), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple* (pp. 320–341). Oxford University Press.

Epstein, R., Lee, V., Mohr, R., & Zankich, V. R. (2022). The Answer Bot Effect (ABE): A powerful new form of influence made possible by intelligent personal assistants and search engines. *PLOS ONE*, *17*(6), e0268081. https://doi.org/10.1371/journal.pone.0268081

Epstein, R., & Robertson, R. E. (2015). The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, *112*(33), E4512–E4521. https://doi.org/10.1073/pnas.1419828112

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298–320. https://doi.org/10.1093/poq/nfw006

Fletcher, R., & Nielsen, R. K. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, *20*(7), 2450–2468. https://doi.org/10.1177/1461444817724170

Garrett, R. K. (2009). Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication*, *59*(4), 676–699. https://doi.org/10.1111/j.1460-2466.2009.01452.x

Gleason, J., Hu, D., Robertson, R. E., & Wilson, C. (2023). Google the gatekeeper: How search components affect clicks and attention.

Proceedings of the International AAAI Conference on Web and Social Media, *17*(1), 245–256. https://ojs.aaai.org/index.php/ICWSM/article/view/22142

Google. (2023). *Ranking results—How google search works*. Retrieved May 3, 2023, from https://www.google.com/search/howsearchworks/how-search-works/ranking-results/

Google. (2024). *Your choice for cookies & data*. Retrieved July 19, 2024, from https://support.google.com/accounts/answer/12065353

Haim, M. (2020). Agent-based testing: An Automated approach toward artificial reactions to human behavior. *Journalism Studies*, *21*(7), 895–911. https://doi.org/10.1080/1461670X.2019.1702892

Haim, M., Arendt, F., & Scherr, S. (2017). Abyss or shelter? On the relevance of web search engines' search results when people google for suicide. *Health Communication*, *32*(2), 253–258. https://doi.org/10.1080/10410236.2015.1113484

Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the filter bubble?: Effects of personalization on the diversity of *Google News*. *Digital Journalism*, *6*(3), 330–343. https://doi.org/10.1080/21670811.2017.1338145

Haim, M., Scherr, S., & Arendt, F. (2021). How search engines may help reduce drug-related suicides. *Drug and Alcohol Dependence*, *226*, 108874. https://doi.org/10.1016/j.drugalcdep.2021.108874

Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring personalization of web search. *Proceedings of the 22nd international conference on World Wide Web*, 527–538. https://doi.org/10.1145/2488388.2488435

Haroon, M., Wojcieszak, M., Chhabra, A., Liu, X., Mohapatra, P., & Shafiq, Z. (2023). Auditing YouTube's recommendation system for ideologically congenial, extreme, and problematic recommendations. *Proceedings of the National Academy of Sciences*, *120*(50), e2213020120. https://doi.org/10.1073/pnas.2213020120

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012. https://doi.org/10.1080/21670811.2019.1623700

Hu, D., Jiang, S., E. Robertson, R., & Wilson, C. (2019). Auditing the partisanship of google search snippets. *The World Wide Web Conference*, 693–704. https://doi.org/10.1145/3308558.3313654

Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., & Mislove, A. (2015). Location, location, location: The impact of geolocation on web search personalization. *Proceedings of the 2015 Internet Measurement Conference*, 121–127. https://doi.org/10.1145/2815675.2815714

Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2015). Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. *Journal of Computer-Mediated Communication*, *20*(2), 171–187. https://doi.org/10.1111/jcc4.12105

Le, H., Maragh, R., Ekdale, B., High, A., Havens, T., & Shafiq, Z. (2019). Measuring political personalization of google news search. *The World Wide Web Conference*, 2957–2963. https://doi.org/10.1145/3308558.3313682

Loecherbach, F. (2023). *Diversity of news consumption in a digital information environment* (Doctoral dissertation). Vrije Universiteit Amsterdam.

Makhortykh, M., Urman, A., & Ulloa, R. (2020). How search engines disseminate information about COVID-19 and why they should do better. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-017

Menchen-Trevino, E., Struett, T., Weeks, B. E., & Wojcieszak, M. (2023). Searching for politics: Using real-world web search behavior and surveys to see political information searching in context. *The Information Society*, *39*(2), 98–111. https://doi.org/10.1080/01972243.2022.2152915

Nechushtai, E., Zamith, R., & Lewis, S. C. (2023). More of the same? Homogenization in news recommendations when users search on Google, YouTube, Facebook, and Twitter. *Mass Communication and Society*, 1–27. https://doi.org/10.1080/15205436.2023.2173609

Oliveira, B., & Lopes, C. T. (2023). The evolution of web search user interfaces—An Archaeological analysis of google search engine result pages. *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, 55–68. https://doi.org/10.1145/3576840.3578320

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, *12*(3), 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*. Penguin UK.

Puschmann, C. (2019). Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism*, *7*(6), 824–843. https://doi.org/10.1080/21670811.2018.1539626

Robertson, R. E., Green, J., Ruck, D. J., Ognyanova, K., Wilson, C., & Lazer, D. (2023). Users choose to engage with more partisan news than they are exposed to on Google Search. *Nature*, *618*(7964), 324–348. https://doi.org/10.1038/s41586-023-06078-5

Robertson, R. E., Lazer, D., & Wilson, C. (2018). Auditing the personalization and composition of politically-related search engine results pages. *Proceedings of the 2018 World Wide Web Conference on World Wide Web—WWW '18*, 955–965. https://doi.org/10.1145/3178876.3186143

Robertson, R. E., & Wilson, C. (2020). Websearcher: Tools for auditing web search. *Proceedings of the 2020 Computation+ Journalism Symposium (Boston, MA, USA)(C+ J 2020)*.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *"Data and Discrimination: Converting Critical Concerns into Productive Inquiry," a preconference at the 64th Annual Meeting of the International Communication Association*, 1–23.

Schwabl, P., Haim, M., & Unkel, J. (2024). Aligning agent-based testing (ABT) with the experimental research paradigm: A literature review and best practices. *Journal of Computational Social Science*. https://doi.org/10.1007/s42001-024-00283-6

Slechten, L., Courtois, C., Coenen, L., & Zaman, B. (2022). Adapting the selective exposure perspective to algorithmically governed platforms: The case of Google Search. *Communication Research*, *49*(8), 1039–1065. https://doi.org/10.1177/00936502211012154

Steiner, M., Magin, M., Stark, B., & Geiß, S. (2022). Seek and you shall find? A content analysis on the diversity of five search engines' results on political queries. *Information, Communication & Society*, *25*(2), 217–241. https://doi.org/10.1080/1369118X.2020.1776367

Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, *30*(3), 341–366. https://doi.org/10.1007/s11109-007-9050-9

Sunstein, C. R. (2001). *Republic.com*. Princeton University Press.

Thurman, N., & Schifferes, S. (2012). The Future of personalization at news websites. *Journalism Studies*, *13*(5–6), 775–790. https://doi.org/10.1080/1461670X.2012.664341

Trielli, D., & Diakopoulos, N. (2019). Search as news curator: The role of google in shaping attention to news information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3290605.3300683

Trielli, D., & Diakopoulos, N. (2022). Partisan search behavior and Google results in the 2018 U.S. midterm elections. *Information, Communication & Society*, *25*(1), 145–161. https://doi.org/10.1080/1369118X.2020.1764605

Ulloa, R., & Kacperski, C. S. (2023). Search engine effects on news consumption: Ranking and representativeness outweigh familiarity in news selection. *New Media & Society*. https://doi.org/10.1177/14614448231154926

Unkel, J., & Haim, M. (2021). Googling politics: Parties, sources, and issue ownerships on google in the 2017 German federal election campaign. *Social Science Computer Review*, *39*(5), 844–861. https://doi.org/10.1177/0894439319881634

Urman, A., & Makhortykh, M. (2023). You are how (and where) you search? Comparative analysis of web search behavior using web tracking data. *Journal of Computational Social Science*, *6*(2), 741–756. https://doi.org/10.1007/s42001-023-00208-9

Urman, A., Makhortykh, M., & Ulloa, R. (2022). The matter of chance: Auditing web search results related to the 2020 U.S. presidential primary elections across six search engines. *Social Science Computer Review*, *40*(5), 1323–1339. https://doi.org/10.1177/08944393211006863

van Hoof, M., Meppelink, C. S., Moeller, J., & Trilling, D. (2024). Searching differently? How political attitudes impact search queries about political issues. *New Media & Society*, *26*(7), 3728–3750. https://doi.org/10.1177/14614448221104405

Webber, W., Moffat, A., & Zobel, J. (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, *28* (4), 1–38. https://doi.org/10.1145/1852102.1852106

Westerwick, A., Kleinman, S. B., & Knobloch-Westerwick, S. (2013). Turn a blind eye if you care: Impacts of attitude consistency, importance, and credibility on seeking of political information and implications for attitudes. *Journal of Communication*, *63*(3), 432–453. https://doi.org/10.1111/jcom.12028

Wojcieszak, M., Menchen-Trevino, E., Goncalves, J. F. F., & Weeks, B. (2022). Avenues to news and diverse news exposure online: comparing direct navigation, social media, news aggregators, search queries, and article hyperlinks. *The International Journal of Press/Politics*, *27*(4), 860–886. https://doi.org/10.1177/19401612211009160

Wonneberger, A., Meijers, M. H. C., & Schuck, A. R. T. (2020). Shifting public engagement: How media coverage of climate change conferences affects climate change audience segments. *Public Understanding of Science*, *29*(2), 176–193. https://doi.org/10.1177/0963662519886474

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., De Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1). https://policyreview.info/articles/analysis/should-we-worry-about-filter-bubbles

Journal of Computer-Mediated Communication, 2024, 29, 1–14
https://doi.org/10.1093/jcmc/zmae020
Full-length Research Article