[1]MANUSCRIPT IN PREPARATION

Searching for BLM:

Search Behavior and Google Results During the 2020 Black Lives Matter Protests

Abstract

The murder of George Floyd in May 2020 ignited waves of nation-wide protests and polarizing discourse on racial justice and the Black Lives Matter (BLM) movement. This study examines online information-seeking tendencies and information exposure in the US in this context using a mixed method approach, including public opinion survey, Google Trends data, and search engine scraping. Survey data reveals that individuals reported search queries that reflected their pro- and anti-attitudes towards the movement. The analysis of aggregate search data indicates that higher search activity for BLM-friendly queries and topics was found in states with higher BLM-support. Last, the analysis of search engine results returned from queries provided by pro- and anti-BLM groups shows evidence of left-leaning bias at the web domain level and slightly different topical focus at the content level. These findings suggest how bias in online search behavior related to BLM results in algorithmic promotion of content that reinforces, rather than challenges, pre-existing viewpoints.

Keywords: online search behavior, public opinion, digital trace, search engines, algorithms, news diversity, search results, Black Lives Matter.

[1] Pleases cite as: Tong, C. & Diakopoulos, N. (2021). Searching for BLM: Search behavior and Google results during the 2020 Black Lives Matter protests. *Paper presented at the 2021 International Communication Association virtual conference, Communication and Technology Division.*

Searching for BLM:

Search Behavior and Google Results During the 2020 Black Lives Matter Protests

Digital platform companies are taking increasingly important roles in the contemporary information life of citizens. People's experience of news is being re-shaped by technological changes, from the ability to get news on mobile devices ('portable'), contribute to news production and transmission ('participatory'), or get 'social' with news consumption, to the ability to customize news to personal interests. Together with social media and their accumulating democratic impact, Internet search engines like Google and Bing are vital information intermediaries that could enable, channel or inhibit exposure to diverse media. For these important implications, the effects of personalization algorithms and search engines on information have recently begun to receive substantive research attention and scrutiny.

The existing literature on 'filter bubbles' demonstrates how personalization technologies customize information to users based on algorithmic predictions about their interests and past behaviors (Möller et al., 2018; Nechushtai and Lewis, 2019). However, there is relatively less attention paid to the effects of human input in the personalization outcomes. Due to the heterogeneity in individuals' information seeking habits, only a few studies have looked at user-generated search queries and considered these as inputs to examine how Google search results may vary as a result. Some past research showed that people take cues from the language used in the media they consume and often employ biased search queries to seek information, which results in further political bias in the results returned (Le et al., 2019; Mustafaraj et al., 2020; Tripodi, 2018). Some others found that differences in political search queries did not produce substantial differences in search results, instead reflecting more of a mainstreaming effect (Trielli and Diakopoulos, 2019, 2020).

The purpose of this study is to extend the line of research on the search queries chosen by people to seek information, and focuses on the two main questions: whether and to what extent people use biased search queries to seek political information? and to what extent are Google search results personalized based on the biased slant of the search terms provided by the users?

To achieve these purposes, we first use a survey to solicit search terms from respondents regarding a current issue, here focusing on the Black Lives Matter (BLM) protests. Compared to long-standing polarized issues in American politics like abortion, racial relations have consistently been on the public agenda and recently gained enormous public attention due to death of George Floyd under police force. The BLM movement is a partisan issue, as a longitudinal survey conducted by CIVIQS on American registered voters since 2017 revealed that 87% of Democrats indicated 'support' towards BLM compared to only 6% of Republicans. In order to understand whether and to what extent this partisanship is expressed in the choice of search queries, the search terms we collected via survey were analyzed to examine how pre-existing opinion on BLM correlates with search query choice. Google Trends data was used to corroborate findings regarding these solicited search queries from respondents. In order to address our second driving question, selected queries were then used to query the Google Search engine and examine how information about BLM was returned in the search results, and how the results could be personalized with different queries reflecting bias.

Findings from the study contribute to understanding how user information seeking tendencies specific to search engines relate to bias in search results and the extent to which results based on biased search queries further reinforces or undermines partisan bias. The results are discussed in terms of the interactions between human action and machine algorithms within the context of the current media system and information diversity.

Literature Review

Search interest as a measure of public attention and opinion

The shift to digital news has paved ways for the emergence of new information habits: people relying on online search engines as much as, if not more than, going to particular news outlets online. Studies have shown that search engines are crucial entry points to many news websites today(Fletcher and Nielsen, 2017; Newman et al., 2020; Olmstead et al., 2011), as roughly 93% of web traffic comes from search engines (Schumacher, 2020). The role of search engines in shaping political opinions has become one of the most important and timely issues in the field of political communication and technology (Epstein and Robertson, 2015; Hargittai, 2007).

Information seeking behaviors measured as trends in Internet search activity are often seen as an indicator of public attention. Research that examined public attentiveness and interest using search trends often emphasized that while search trends are reflective of public attention, they do not necessarily indicate individual or aggregate attitudes about certain topics (Ripberger, 2011). Public attention was measured by metrics such as media coverage, whereas public opinion was measured by metrics such as opinion polls (Jungherr et al., 2017; McCombs and Valenzuela, 2020). While it can be dangerous to conflate the two potentially distinct concepts, the relationship between opinion and attention is an empirical question subject to verification when information seeking at both individual and aggregate levels are concerned. Surveys, especially when conducted in a timely manner, can be used to tap into individuals' thought process and information seeking behavior. This is informative and crucial in understanding public opinion as individuals approach online information seeking with diverse goals and motivations. Aggregate search behaviors, on the other hand, reveal the dynamic and ephemeral nature of public attention. In combination, these two measures can provide better insights into the attitudinal and behavioral processes behind search activities and attention to public issues. As later illustrated in our findings, to the extent that search intention at the individual level is an indication of a citizen's interest and attention, and is motivated by some degree of thought and deliberation, the way people report their search interests in BLM could reveal their attitudes and opinions on the topic.

User bias reflected in individual search behaviors

Previous research has well documented the innate biased nature of humans and human behaviors. The literature on 'selective exposure', for example, notes the general tendency of individuals to prefer attitude-consistent information over attitude-discrepant ones (Festinger, 1957; Flaxman et al., 2016; Stroud, 2017). Politically, audiences are increasingly fragmented by partisanship and ideology fueled by the rise of markedly partisan media to cater to their partisan information needs (Abramowitz and Saunders, 2006; Webster and Abramowitz, 2017).

Research in psychology and information retrieval documents several cognitive biases associated with information seeking. These include prior beliefs about the search topic, the tendency to conduct searches to verify such beliefs and to interpret search results in a way that supports the original idea ('confirmation bias') (Fiske and Taylor, 2013; Knobloch-Westerwick et al., 2015; Knobloch-Westerwick and Kleinman, 2011; White and Horvitz, 2015).

From a technical perspective, human-sourced bias can be found in all processes of algorithmic design and functioning (Bozdag, 2013), in particular, the way people interact with the system (Friedman and Nissenbaum, 1996), which could foster personalization algorithms that amplify confirmation bias and increase informational fragmentation. In the extreme scenarios, this leads to information diets completely devoid of heterogeneity and attitude-countering perspectives ('filter bubbles') (Pariser, 2011).

Empirical and anecdotal evidence shows that search activities performed by humans are biased. For example, one third of the search phrases that users revealed to have performed in the period leading up to the 2018 elections included either semantic bias (i.e., including language that denotes bias independently of the context such as *'the best candidate'* or *'will Beto win'),* or pragmatic bias (i.e., indicating bias as a result of its context within a broader narrative, for example *'blue wave'* or *'Diane Feinstein's age')* (Mustafaraj et al., 2020). Algorithms can also augment these biases and exacerbate the issue via the suggestions they make based on what users type into the search bar (Olteanu et al., 2020; Robertson, Jiang, et al., 2018).

Search queries put into the search bar could also be an artifact of consuming ideologically biased news. Evidences of selective exposure indicates that individuals do not consume pro- and counter- attitudinal information equally (Knobloch-Westerwick and Kleinman, 2011; Muddiman and Stroud, 2017). Ethnographic work intriguingly shows that this tendency to prefer congenial information is associated with the tendency to verify information using words or phrases from exactly the same biased media sources (Tripodi, 2018). In particular, conservative Republicans were found to use Google to 'do their own research' but doing so by using the exact phrases delivered to them by the conservative media they consume, which leads to further conservative bias in the returned results.

Biased search queries can come from search engine algorithms as well. Research on media manipulation specifies that search engines are often weaponized by disinformation actors to mislead and influence public perceptions. This is done by the strategic creation and amplification of problematic search queries (e.g., '*did the Holocaust happen*') that can lead people to entirely different spheres of information depending on the queries they use (Golebiewski and boyd, 2018). Moreover, politically slanted search terms were found to lead to results that further tailor to preexisting political predispositions (Le et al., 2019). For example, search terms such as *'carbon footprint', 'comprehensive immigration reform', 'Paris climate agreement', 'support our veterans',* and *'uninsured Americans'* reinforce the liberal search results, whereas *'flat tax', 'Medicare for all',* and *'national debt'* lead to more conservative personalized search results.

Past studies highlighted the role of user-input biases in influencing how a dynamic curation algorithm responds to an individual. In the context of search engines, this refers to the agency when users trigger search engines using certain search queries. For example, Trielli and Diakopoulos (2020) showed that search engine users of different political ideologies differ in the way they search for political information about political candidates, measured by the search terms they used.

In short, the literature demonstrates that in general, as individuals often seek for information with particular motives and goals, bias is inadvertently manifested in the act of acquiring information. In particular, as members of the public constantly receive cues and heuristics from surrounding information sources to validate and seek further information, this introduces bias in the form of search queries, which has the potential of prompting search engines to return results relevant to these specific queries. We form the following RQ and corresponding hypotheses:

RQ1: Is there a pattern of selective use of search queries that reflect political predispositions regarding attitudes towards BLM?

H1a: At the individual level, people will report search terms that reflect their opinions about BLM.

H1b: At the aggregate level, average search interest will reflect political reality surrounding public support for BLM.

Search engine bias and effects on source and information diversity

The science built into search engine algorithms was first derived from social science, the idea of the structures of citations (Brin and Page, 1998). In the past, what was determined to be 'relevant' by search engines was the match between subject matter (query) and document (page) and results were ranked based on the number of influential in-links to a site. Overtime, search engines' definition of 'relevance' increasingly takes into account users' interaction (i.e., clicks) in determining what are considered positive feedback (Bozdag, 2013). Considering its dominant market share status, the Google Search engine has attracted relatively more research focus than other web search engines.

There have been several critiques that are specifically related to search results bias. In particular, search engines are often criticized for amplifying 'the rich get richer' effect, as results show a great unequal distribution in terms of web domains at the top of the search returns, in which the top 20% (often influential and highly authoritative sources) are prioritized. This critique is rooted from an idealistic notion of a democratizing and equal web sphere and how search engines undermine such vision. The built-in mechanism of search engines that boost sites with the most links from other sites (which themselves also receive a lot of links) perpetuates the promotion of the socially privileged (Hindman, 2008; Introna and Nissenbaum, 2000) and arguably undermines information diversity.

Other studies focused on the bias of personalization algorithms, in particular how search personalization influences the results (Hannak et al., 2013; Kliman-Silver et al., 2015; Robertson, Lazer, et al., 2018) or the political bias in search results (Diakopoulos et al., 2015; Hu et al., 2019; Robertson, Jiang, et al., 2018).

The literature suggests that search personalization is driven most strongly by the account log-in status and geolocation (Hannak et al., 2013; Kliman-Silver et al., 2015; Robertson, Lazer, et al., 2018). Other endogenous factors include the root query, language settings, Web history, clicking behavior (Ørmen, 2015), and multiple use of search engine services (e.g. Google Drive, Google Plus, etc. of Google) (Robertson, Lazer, et al., 2018). Exogenous variables such as A/B testing, experimentation and randomization can also be present, introducing result variation even for the same query and same person (Diakopoulos et al., 2018).

Regarding news and political domains, research found very minimal personalization effects that support the ideological filter bubble thesis (Haim et al., 2018; Nechushtai and Lewis, 2019; Puschmann, 2018; Trielli and Diakopoulos, 2020). For example, Nechushtai & Lewis (2019) using search result data from actual users showed that users with different political leanings from different locations saw very similar Google news recommendations about Hillary Clinton and Donald Trump, and the top recommendations were consistently identical for conservatives and liberals. On the other hand, the extent to which Google maintains separate partisan narratives and reinforces ideological filter bubbles by its design is documented anecdotally and empirically. For example, services like Google and YouTube have been found to lead users down the rabbit hole of extremist content through algorithmic recommendations (e.g. Noble, 2018). Political actors with a political agenda in mind also take advantage of search engines by using strategic terms and problematic queries to lead people to inaccurate or disturbing information (Golebiewski & boyd, 2019). Insights from the literature suggests the importance of examining the connections search engine effects and partisan politics related to the BLM movement. Thus, we form the following research question:

RQ2: How do user-generated queries regarding BLM relate to search results returned from these queries?

H2a. At the web source level, queries of contrasting slants return different sets of sources in Google Search results.

H2b. At the content level, queries of contrasting slants return contents with differing topical prevalence.

Methods

Survey sample

The death of George Floyd under police force on 25 May 2020 triggered a series of nation-wide protests filled with outrage at the atrocity of the acts of the police officers involved. We recruited a sample of 511 respondents from the US via panels provided by Qualtrics shortly after the peak of the protests (4 – 30 August 2020) to collect information about their media use and attitudes towards the BLM movement and the ongoing protests.

Measures of user-generated BLM search queries

To capture search interest, we included an open-ended question asking respondents for 3 search terms regarding the recent BLM protests. The exact wording for the question is: *'We are trying to understand how people seek for information regarding current news events. If you were to conduct an online search to get information about the recent protests, what terms would you search for?'*

In terms of attitudes towards BLM protests, support or non-support towards BLM protests was measured with respondents' self-reported agreement to this statement: 'I support the BLM protests' (0: strongly disagree to 5: strongly agree) (Mean = 2.75, Median = 3, SD = 1.96). Based on the bimodal distribution of this variable, the data was split by the Median: People scoring 0, 1, 2 are counted as 'BLM Opposers' (N = 218), and those scoring 3, 4, 5 are counted as 'BLM Supporters' (N = 293).

Respondents' answers were then coded and grouped into two groups of search queries based on their attitudes towards the BLM protests. After cleaning the data with the OpenRefine tool and removing responses containing gibberish or nonsensical answers, the data suitable for analysis includes 1172 search queries (484 from the 'BLM Opposers' group and 688 from the 'BLM Supporters' group) (See Table A1 in the Appendix for major themes).

Text mining techniques were used to compare the text corpus comprised of search queries from the two groups (Silge and Robinson, 2017). The cleaning process involved tokenizing the corpus into n-grams (unigrams and bigrams), removing common English stop words and calculating n-gram frequencies for each query group.

To compare the language usage of BLM supporters and opposers and to have more robust statistical measures to determine if the observed n-grams have a higher likelihood of being used by both groups versus by one group more than the other, the log likelihood ratio test was used to determine how an n-gram was more or less likely to come from either group. A positive log odds ratio implies greater usage amongst opposers and a negative log odds ratio a greater usage amongst supporters.

Google Search trends data

Search interest in topics related to the BLM movement was downloaded from Google Trends. Google Trends has been widely used to understand how aggregate search behavior can inform researchers on public interest on particular topics[2]. Past research has used Google search data to explore the spread of diseases (Eysenbach, 2011; Ginsberg et al., 2009; Hulth and Rydevik, 2011), or to understand the kind of searches people perform to find information important to them (Fallows, 2005; Pew Research Center, 2019; Trevisan et al., 2018). Data from Google Trends represents proportional percentages of searches made for a certain query out of all searches made in a particular place and time period.

We use Google Trends to obtain search data for several queries that were supplied by the respondents. In particular, the relative search volume for these queries in each of 50 U.S. states from the date range of 25 May – 31 August 2020 were collected. Search trends data were then combined with CIVIQS data also broken down by states, which included 193,405 responses (April 25, 2017 – October 26, 2020) on public support or opposition towards the BLM movement[3]. If a state leans more towards 'Support' for BLM (according to CIVIQS data), search volume for queries indicating 'support' should be relatively higher than that for queries indicating 'oppose' within that state. In other words, search volume for 'support' queries should be positively correlated with the level of BLM support and negatively correlated with the level of BLM opposition in each state and vice versa.

Google Search results

---

[2] https://support.google.com/trends/answer/4365533

[3] CIVIQS is a research company that tracks public opinion on a range of political issues by conducting daily interviews with registered voters. To produce daily measures of public opinion, a statistical tracking model is used to correct for random differences between the demographics of the sample and the population, and smooth out day-to-day sampling variation. More information on the detailed methodology can be found here.

From the list of queries, an equal number of queries was taken from the two groups: BLM Supporters and BLM Opposers (N = 17 queries per group). These queries are selected based on their informativeness as search queries. The majority of the open-ended responses from the respondents were not detailed enough beyond iterations of 'black lives matter' or 'black lives matter protests'. This is reflective of the actual information-searching behaviors in the real-world. For our purpose of investigating search results from both common and specific queries, we sampled 34 queries indicating support and opposition towards BLM (17 each) and used them to automatically query Google Search (The full list is included in the Appendix). The returned URLs in the search engine result pages (SERPs) were scraped using open software (Vincent, 2020)[4] and the domain block of text from each URL was extracted for further analysis. The process of automatically querying Google Search resulted in non-personalized search results (with no user data stored such as browsing or searching histories, Internet cookies, etc.), from only a single location (a mid-western state).

Results

Differences in search queries as a function of attitudes towards BLM

H1 posits that people will report search terms regarding BLM that reflect their beliefs and value judgments. The analysis of user-generated search queries from survey respondents found some support for the hypothesis. Words like 'black', 'defund', 'affairs' were frequently found in both sets of search terms by BLM supporters and opposers, whereas words like 'equality', 'brutality' and 'unrest' showed up more in the BLM supporters' set and 'Antifa', 'Fox', 'riots' showed up more in the BLM opposers' set.

[4] https://github.com/nickmvincent/LinkCoordMin

To give more context and clarity into these word selections, bigrams were included. Table 1 displays the unigrams and bigrams from the search queries that were roughly equally likely to come from both groups. Accordingly, both BLM supporters and opposers were about equally likely to report search terms about ‘black lives’, ‘Chicago protests’, or ‘news’.

Figure 1 illustrates the probabilities of n-gram occurrences and distinct language use in each group. ‘George Floyd’, ‘equality’, ‘police brutality’, ‘racial injustice’, etc. were more likely to come from the supporters of the BLM movement. In contrast, ‘Antifa’, ‘looting’, ‘private property’ and ‘protester violence’ were seen occurring more often in the search queries provided by BLM opposers. Put together, these findings indicated that search queries regarding BLM solicited from individuals reflect biases aligned with their support or opposition to BLM, even with neutral question wording.

[TABLE 1 AND FIGURE 1 ABOUT HERE]

Correlation of aggregate search interests and BLM support level

Data from CIVIQS and Google Trends was used to assess whether biased search terms from the respondents in our survey might be correlated to general public support or opposition towards BLM. Table 2 presents the correlations between Google Trends of ‘support’ and ‘oppose’ queries and the percentage of BLM support/opposition by state (i.e. CIVIQS data). As we expected, the percent of state support for BLM was positively correlated with search interest in ‘support’ queries and negatively correlated with search interest in ‘oppose’ queries (though the negative correlations with some ‘oppose’ queries was somewhat weaker); whereas the opposite patterns were found for state opposition of BLM. Overall, this indicates that across 50 states in the U.S., search interest for support or oppose queries as identified in our survey sample generally reflected levels of support or opposition towards BLM as measured by CIVIQS data.

[TABLE 2 ABOUT HERE]

Top domains and political bias analysis

The organic search results (i.e., the blue links) returned from the 'support' and 'oppose' queries were analyzed (N = 168 URLs per query group)[5]. The web domains of each link were extracted, and the top domains are shown in Table 3. Comparison of the top domains indicates that there was some degree of overlap between the two domain lists, with five domains being commonly shared among the top domains (*Wikipedia.org, cnn.com, washingtonpost.com, nytimes.com, theguardian.com*). Domains such as *theatlantic.com*, *latimes.com*, *vox.com* and *npr.com* were significantly more likely to appear in the 'oppose' queries SERPs, whereas *aclu.org, usatoday.com, and fivethirtyeight.com* occurred relatively more in the 'support' queries SERPs. This suggests that Google presents users with a largely overlapping set of sources of information, as well as some diverging sources based on the biased nature of the queries.

[TABLE 3 ABOUT HERE]

Each domain from the lists of URLs in the SERPs for 'support' and 'oppose' queries was assigned a partisan bias score to determine the extent of political bias in results regarding BLM. These bias scores were derived from the sharing patterns of web domains by more than 500,000 known Democrats and Republicans on Twitter, ranging from -1 (shared only by Democrats) to 1 (shared only by Republicans) (Robertson, Jiang, et al., 2018). As such these scores reflect the biases of the people sharing the domain, which is important to consider when interpreting these results. Figure 2 displays the distribution of the biases for the organic links in SERPs for the support-queries and oppose-queries. As can be seen, both distributions are skewed

[5] The majority of queries (32 out of 34) returned 10 organic links. The two queries 'systematic racism' and 'tearing down of historical statues' return 8 organic links in the first SERP.

slightly left towards negative scores, indicating that both sets of queries returned more URLs shared by liberally biased users. The average bias scores suggested that the support-URLs (mean = -0.329, SD = 0.285) were more liberally biased than the oppose-URLs (mean = -0.234, SD = 0.322). Results from the Kolmogorov-Smirnov test indicated statistical difference in the distributions of the two groups (0.185, p-value = 0.007). Moreover, there is a larger spread in scores for the oppose queries: at the extreme positive end of the distribution (i.e. shared by Republicans) we see that no scores above 0.5 were observed for support queries, whereas some scores above 0.5 were observed for oppose queries. We conducted further validation of the result with other domain bias classifications, including the Facebook audience-based scores (Bakshy et al., 2015) and AllSides.com crowdsourced rating system, and found that this observation of the liberal-leaning bias holds across the three metrics (see the Appendix).

[FIGURE 2 ABOUT HERE]

Content analysis with structural topic modelling

We used Structural Topic Modelling (STM) to identify the contents of these results and to examine whether 'support' and 'oppose' queries generate results that focus on different aspects regarding the BLM movement. STM is a computer-assisted textual analysis technique built on the tradition of probabilistic topic models such as Latent Dirichlet Allocation (LDA), which allows not only the detection of topics in text corpora but also the estimation of topics' relationship to document metadata (Roberts et al., 2014). For our purpose, STM was used to reveal the topical contents of BLM discourse, and statistically examine if topical focus would differ depending on the 'support' or 'oppose' slant of search queries.

STM revealed ten topics in the text content of search results returned from both types of queries. We named these topics according to: (1) words with the highest probability within each topic, (2) words that are both frequent and exclusive to each topic, (3) the log probabilities of particular words conditional on each topic ('score') (Chang, 2015), and (4) the exemplary texts of each topic (Roberts et al., 2014). Lastly, to verify these labels and inspect nuanced differences among these topics, manual validation was conducted on random documents for each topic. Table 4 illustrate the STM topics, topic proportions and top terms in each topic.

[TABLE 4 ABOUT HERE]

We expected that topical prominence would differ based on the Support/Oppose slants of the queries. Table 5 shows the ordinal logistic regression results with query slant as the predictor and topic probability as the dependent variable. Accordingly, of the ten topics, Support queries resulted in more content about racial disparities (Topic 5, $\beta = 0.048$, $p < .05$) and donation/bailout (Topic 9, $\beta = 0.121$, $p < .001$), whereas Oppose queries resulted in more content about looting/damage (Topic 1, $\beta = -0.122$, $p < .001$), and race and politics (Topic 3, $\beta = -0.052$, $p < .05$). These suggest that using either support or oppose queries might lead users to somewhat differing discourse about the BLM protests.

[TABLE 5 AND FIGURE 3 ABOUT HERE]

Discussion

The goal of this study is to understand how user bias in information-seeking tendencies relates to bias in search engine results and the extent to which search results reflect such biases. We found evidences from survey data that illustrated how preexisting opinions on BLM colored the way respondents reported their intended search terms regarding the topic. Aggregate search data gave credence to these findings regarding how search volumes for biased queries correspond to the broad level of support or opposition towards the BLM movement in each state. Finally, the analysis of web domains in SERPs returned from users-supplied search terms and topical distribution of contents showed that the choice of queries, in aggregate, will lead to sources that generally reflect that political position.

Regarding information-seeking at the individual level, the findings are closely in line with previous research on congruent informational selectivity depending on one's preexisting attitude on the issue (Iyengar et al., 2008). From the goals and motivations approached, individuals are not passive information seekers; they can be goal-oriented and active in seeking out specific types of information. However, under conditions of no or little motivation, when being asked (with very neutral question wording) to indicate information-seeking intentions on a topic, this study once again illustrated how people gave spontaneous query choices that reflected predilections for familiar and attitude-congruent information. The widespread lack of sophistication among the public in using search engine technology (Hindman, 2008), including the use of short, generic search terms and reliance on known sources, may limit the ability to locate political information that widens and challenges viewpoints. Since search engines are a user-centric tool, and users are broadly diverse in skills and motivations (Hargittai, 2010; Klawitter and Hargittai, 2018), a direction for future research would be to examine the conditions under which biased searching can be induced or suppressed. For instance, a related and extended research question is whether the tendency to supply attitudinally congruent search queries still holds when survey questions are asked or framed differently.

Regarding the search results, the findings indicated a strong mainstreaming effect in the top URLs for user-generated biased queries. Across both sets of queries with contrasting slants towards BLM, search results were dominated by mainstream news sites, politically related domains and Wikipedia. Importantly, both sets of queries generated more liberally biased URLs (with the 'support' queries URLs expectedly more liberally biased than the 'oppose' queries

URLs), which resulted in an overall skew towards the left of the political spectrum. This observation is consistent with arguments regarding the overall left skew of the mainstream media system (Gentzkow and Shapiro, 2010; Groseclose, 2011; Ladd, 2011). Relatedly, the findings also suggest that to some extent, the partisan reinforcing effects of search results are asymmetric, with the conservative-leaning queries more likely to lead to sources that undermine such bias. For instance, 'Oppose' queries regarding BLM such as '*blue lives better*', '*protesters and crimes*' or '*cities burning*' resulted in domains that were skewed further out on the right distribution (including *bluelivesmatter.blue* (.813), *nationalreview.com* (.636)) but also turned up domains like *aclu.org* (-.791), *thenation.com* (-.730), *newyorker.com* (-.562), or *vox.com* (-.555) which were widely shared by liberal social media users. When looking into the topical contents returned for contrasting query slants (i.e., going past the domain analysis), we found that the query slant somewhat shaped the slant of individual articles surfaced. In line with previous research on how even the same web domains can lead to different content (Rauchfleisch et al., 2020), the findings here indicated that the topical focus of content would be somewhat different depending on the slant (support or oppose) of the queries, which suggests that online users can be exposed to materials that solidify rather than challenging their viewpoints.

The study has several limitations. One was that there was a time lag between the height of the George Floyd protests (end May – June) and the period of survey data collection (in August where protests still remained but had subsided noticeably), which prevented us from accurately capturing public sentiment at its peak, and capturing the SERPs in real-time. Regarding the former, if anything, the results suggest that this could potentially underestimate the actual effect of how public sentiment was translated into search intentions regarding BLM protests. To the degree that search queries mirror positions, search queries solicited during the peak of BLM protests could be even more biased in tone. As to whether the political bias in search results would be different at an earlier timepoint, this remains unanswered, as unfortunately there is no way to comprehensively collect SERPs from the past. Interested readers can find some SERPs stored in versions of Web Archive, but not for very many queries.

Another limitation of the study is that we focused on a single search engine to analyze the potential for information diversity in the context of search engines. Other studies have found variations in search results on socially important topics provided by different search engines (Makhortykh et al., 2020; Urman et al., 2021), particularly the way in which different information sources such as government-related websites or alternative media are prioritized. Comparison across multiple search engines is another fruitful area for future research.

To sum up, this study demonstrates that search queries disclose individual issue positions and can shape or influence the returned results to a certain degree, but this is potentially limited by the supply of mainstream media coverage. With their combined effects, the interactions between human action and machine algorithms could potentially lead people to different spheres of information and limit access to attitudinally challenging points of view. The mainstreaming effect observed in how search engines prioritize authoritative media suggests that individuals with diverging issue positions are exposed to at least some common topics and agendas, despite a slightly different topical focus. Given the importance of search channels as gateways to online information, more empirical investigations are needed to understand the connection between users' choice of search queries and the information they are exposed to in the information-seeking process.

Declaration of interest

No potential conflict of interest was reported by the authors.

References

Abramowitz AI and Saunders KL (2006) Exploring the Bases of Partisanship in the American Electorate: Social Identity vs. Ideology. *Political Research Quarterly* 59(2). SAGE Publications Inc: 175–187. DOI: 10.1177/106591290605900201.

Bakshy E, Messing S and Adamic L (2015) Exposure to Ideologically Diverse News and Opinion on Facebook. *Science* 58: 707.

Bozdag E (2013) Bias in algorithmic filtering and personalization. *Ethics and Information Technology* 15(3): 209–227. DOI: 10.1007/s10676-013-9321-6.

Brin S and Page L (1998) The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30(1). Proceedings of the Seventh International World Wide Web Conference: 107–117. DOI: 10.1016/S0169-7552(98)00110-X.

Chang J (2015) *Lda: Collapsed Gibbs Sampling Methods for Topic Models*. Available at: https://CRAN.R-project.org/package=lda (accessed 6 December 2021).

Diakopoulos N, Trielli D and Mussenden S (2015) Why Google Search Results Favor Democrats. Available at: https://slate.com/technology/2015/12/why-google-search-results-favor-democrats.html (accessed 17 April 2019).

Diakopoulos N, Trielli D, Stark J, et al. (2018) I Vote For— How Search Informs Our Choice of Candidate. In: *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. M. Moore and D. Tambini (Eds.). Available at: https://www.academia.edu/37432634/I_Vote_For_How_Search_Informs_Our_Choice_of_Candidate (accessed 29 April 2021).

Epstein R and Robertson RE (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. DOI: 10.1073/pnas.1419828112.

Eysenbach G (2011) Infodemiology and Infoveillance: Tracking Online Health Information and Cyberbehavior for Public Health. *American Journal of Preventive Medicine* 40(5, Supplement 2). Cyberinfrastructure for Consumer Health: S154–S158. DOI: 10.1016/j.amepre.2011.02.006.

Fallows D (2005) Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve. In: *Pew Internet & American Life Project: Search Engine Users*. Available at: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2005/PIP_Searchengine_users.pdf.pdf.

Festinger L (1957) *A Theory of Cognitive Dissonance*. Stanford University Press.

Fiske ST and Taylor SE (2013) *Social Cognition: From Brains to Culture*. SAGE.

Flaxman S, Goel S and Rao MJ (2016) Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly* 80: 298.

Fletcher R and Nielsen RK (2017) Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. *Journal of Communication* 67(4): 476–498. DOI: 10.1111/jcom.12315.

Friedman B and Nissenbaum H (1996) Bias in computer systems. *ACM Transactions on Information Systems*: 330–347.

Gentzkow M and Shapiro JM (2010) What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica* 78(1): 35–71. DOI: 10.3982/ECTA7195.

Ginsberg J, Mohebbi MH, Patel RS, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–1014.

Golebiewski M and boyd danah (2018) Where Missing Data Can Easily Be Exploited. In: *Data & Society*.

Groseclose T (2011) *Left Turn: How Liberal Media Bias Distorts the American Mind*. St. Martin's Publishing Group.

Haim M, Graefe A and Brosius H-B (2018) Burst of the Filter Bubble?: Effects of personalization on the diversity of Google News. *Digital Journalism* 6(3): 330–343. DOI: 10.1080/21670811.2017.1338145.

Hannak A, Sapiezynski P, Molavi Kakhki A, et al. (2013) Measuring Personalization of Web Search. In: *Proceedings of the 22Nd International Conference on World Wide Web*, New York, NY, USA, 2013, pp. 527–538. WWW '13. ACM. DOI: 10.1145/2488388.2488435.

Hargittai E (2007) The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction.". *Journal of Computer-Mediated Communication* 12: 769.

Hargittai E (2010) Digital Na(t)ives? Variation in Internet Skills and Uses among Members of the "Net Generation"*. *Sociological Inquiry* 80(1): 92–113. DOI: https://doi.org/10.1111/j.1475-682X.2009.00317.x.

Hindman M (2008) *The Myth of Digital Democracy*. Princeton University Press. DOI: 10.1515/9781400837496.

Hu D, Jiang S, Robertson RE, et al. (2019) Auditing the Partisanship of Google Search Snippets. *Proceedings of the 2019 World Wide Web Conference (WWW'19), May 2019, San Francisco, CA, USA*: 12. DOI: 10.1145/3308558.3313654.

Hulth A and Rydevik G (2011) Web query-based surveillance in Sweden during the influenza A(H1N1)2009 pandemic, April 2009 to February 2010. *Eurosurveillance* 16(18).

European Centre for Disease Prevention and Control: 19856. DOI: 10.2807/ese.16.18.19856-en.

Introna D L and Nissenbaum H (2000) Shaping the Web: Why the Politics of Search Engines Matters. *Information Society* 16: 169.

Iyengar S, Hahn KS, Krosnick JA, et al. (2008) Selective Exposure to Campaign Communication: The Role of Anticipated Agreement and Issue Public Membership. *The Journal of Politics* 70(1). The University of Chicago Press: 186–200. DOI: 10.1017/S0022381607080139.

Jungherr A, Schoen H, Posegga O, et al. (2017) Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support. *Social Science Computer Review* 35(3). SAGE Publications Inc: 336–356. DOI: 10.1177/0894439316631043.

Klawitter E and Hargittai E (2018) "It's Like Learning a Whole Other Language": The Role of Algorithmic Skills in the Curation of Creative Goods. *International Journal of Communication* 12(0). 0: 21.

Kliman-Silver C, Hannak A, Lazer D, et al. (2015) Location, Location, Location: The Impact of Geolocation on Web Search Personalization. In: *Proceedings of the 2015 Internet Measurement Conference*, New York, NY, USA, 2015, pp. 121–127. IMC '15. ACM. DOI: 10.1145/2815675.2815714.

Knobloch-Westerwick S and Kleinman SB (2011) Preelection Selective Exposure: Confirmation Bias Versus Informational Utility. *Communication Research*. DOI: 10.1177/0093650211400597 (accessed 20 April 2021).

Knobloch-Westerwick S, Johnson BK and 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. DOI: 10.1111/jcc4.12105.

Ladd JM (2011) *Why Americans Hate the Media and How It Matters*. Princeton University Press. DOI: 10.1515/9781400840359.

Le H, Maragh R, Ekdale B, et al. (2019) Measuring Political Personalization of Google News Search. In: *The World Wide Web Conference on - WWW '19*, San Francisco, CA, USA, 2019, pp. 2957–2963. ACM Press. DOI: 10.1145/3308558.3313682.

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

McCombs M and Valenzuela S (2020) *Setting the Agenda: Mass Media and Public Opinion*. John Wiley & Sons.

Möller J, Trilling D, Helberger N, et al. (2018) Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society* 21(7). Routledge: 959–977. DOI: 10.1080/1369118X.2018.1444076.

Muddiman A and Stroud NJ (2017) News Values, Cognitive Biases, and Partisan Incivility in Comment Sections. *Journal of Communication* 67(4): 586–609. DOI: 10.1111/jcom.12312.

Mustafaraj E, Lurie E and Devine C (2020) The Case for Voter-Centered Audits of Search Engines During Political Elections. *Conference on Fairness, Accountability, and Transparency (FAT*'20), January 27-30, 2020, Barcelona, Spain.*: 11. DOI: https://doi.org/10. 1145/3351095.3372835.

Nechushtai E and Lewis SC (2019) What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior* 90: 298–307. DOI: 10.1016/j.chb.2018.07.043.

Newman N, Fletcher R, Schulz A, et al. (2020) Reuters Institute Digital News Report 2020.

Olmstead K, Mitchell A and Rosenstiel T (2011) Navigating News Online: Where People Go, How They Get There and What Lures Them Away. In: *Pew Research Center's Project for Excellence in Journalism*.

Olteanu A, Diaz F and Kazai G (2020) When Are Search Completion Suggestions Problematic? *Proceedings of the ACM on Human-Computer Interaction* 4(CSCW2): 1–25. DOI: 10.1145/3415242.

Ørmen J (2015) Googling the news: Opportunities and challenges in studying news events through Google Search. *Digital Journalism* 4(1): 107–124.

Pariser E (2011) *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin.

Pew Research Center (2019) Americans and Digital Knowledge. In: *Pew Research Center: Internet, Science & Tech*. Available at: https://www.pewresearch.org/internet/2019/10/09/americans-and-digital-knowledge/ (accessed 19 April 2021).

Puschmann C (2018) Beyond the Bubble: Assessing the Diversity of Political Search Results. *Digital Journalism*: 1–20. DOI: 10.1080/21670811.2018.1539626.

Rauchfleisch A, Vogler D and Eisenegger M (2020) Transnational News Sharing on Social Media: Measuring and Analysing Twitter News Media Repertoires of Domestic and Foreign Audience Communities. *Digital Journalism* 8(9). Routledge: 1206–1230. DOI: 10.1080/21670811.2020.1835511.

Ripberger JT (2011) Capturing Curiosity: Using Internet Search Trends to Measure Public Attentiveness. *Policy Studies Journal* 39(2): 239–259. DOI: 10.1111/j.1541-0072.2011.00406.x.

Roberts ME, Stewart BM, Tingley D, et al. (2014) Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science* 58(4): 1064–1082. DOI: 10.1111/ajps.12103.

Robertson RE, Jiang S, Joseph K, et al. (2018) Auditing Partisan Audience Bias within Google Search. *Proceedings of the ACM on Human-Computer Interaction* 2(CSCW): 1–22. DOI: 10.1145/3274417.

Robertson RE, Lazer D and Wilson C (2018) Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages. In: *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, Lyon, France, 2018, pp. 955–965. ACM Press. DOI: 10.1145/3178876.3186143.

Schumacher D (2020) Search Queries: What is a Search Query + How To Use Them To Your Advantage for SEO. Available at: https://serp.co/seo/search-query/ (accessed 6 December 2021).

Silge J and Robinson D (2017) *Text Mining with R: A Tidy Approach, by Julia Silge and David Robinson.* Cambridge University Press. Available at: https://www.cambridge.org/core/journals/natural-language-engineering/article/abs/text-mining-with-r-a-tidy-approach-by-julia-silge-and-david-robinson-sebastopol-ca-oreilly-media-2017-isbn-9781491981658-xi-184-pages/3CB30936E5E1B8E5D2A1B4AE365BBFB9 (accessed 6 December 2021).

Stroud NJ (2017) Selective Exposure Theories. *The Oxford Handbook of Political Communication*. Oxford University Press. DOI: 10.1093/oxfordhb/9780199793471.013.009_update_001.

Trevisan F, Hoskins A, Oates S, et al. (2018) The Google Voter: Search Engines and Elections in the New Media Ecology.". *Information, Communication & Society* 21(1): 111–128.

Trielli D and Diakopoulos N (2019) Search as News Curator: The Role of Google in Shaping Attention to News Information. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, Glasgow, Scotland Uk, 2019, pp. 1–15. ACM Press. DOI: 10.1145/3290605.3300683.

Trielli D and Diakopoulos N (2020) Partisan search behavior and Google results in the 2018 U.S. midterm elections. *Information, Communication & Society*: 1–17. DOI: 10.1080/1369118X.2020.1764605.

Tripodi F (2018) Searching for Alternative Facts. *2018*: 64. DOI: https://datasociety.net/wp-content/uploads/2018/05/Data_Society_Searching-for-Alternative-Facts.pdf.

Urman A, Makhortykh M and Ulloa R (2021) 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*. SAGE Publications Inc: 08944393211006863. DOI: 10.1177/08944393211006863.

Webster SW and Abramowitz AI (2017) The Ideological Foundations of Affective Polarization in the U.S. Electorate. *American Politics Research* 45(4): 621–647. DOI: 10.1177/1532673X17703132.

White RW and Horvitz E (2015) Belief Dynamics and Biases in Web Search. *ACM Transactions on Information Systems* 33(4): 18:1-18:46. DOI: 10.1145/2746229.

Appendix

A1. Main themes in user-generated search queries

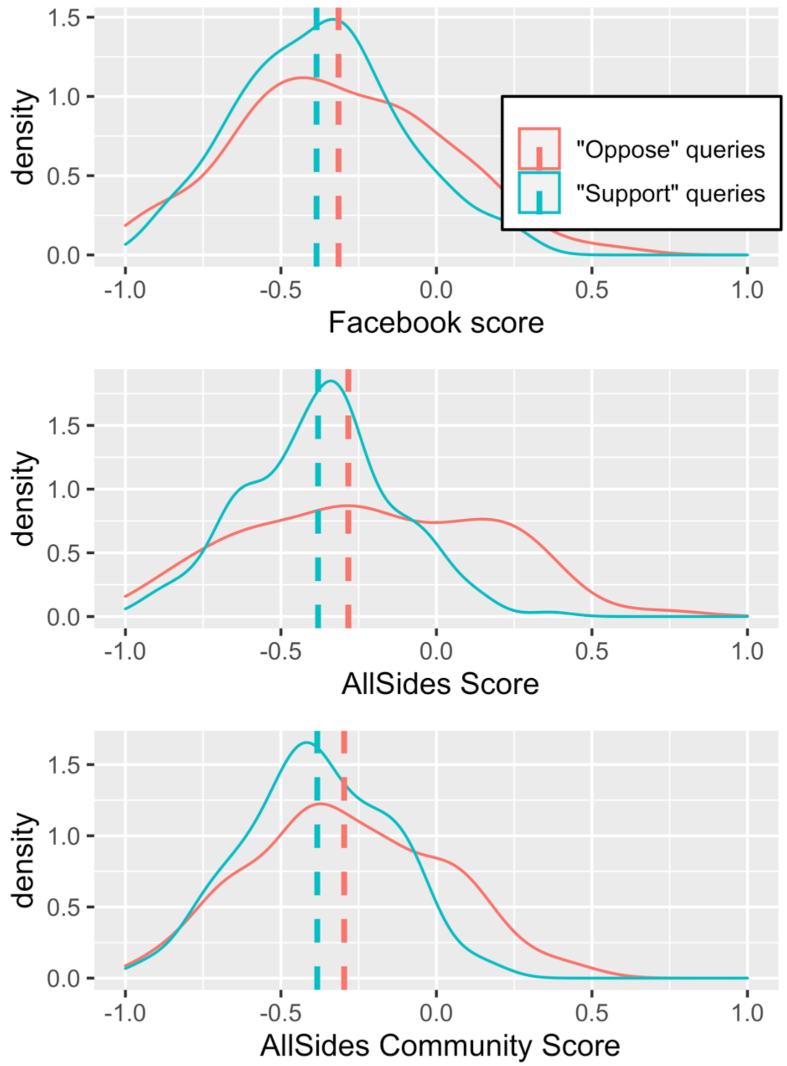
Category	Definition	Examples
Black lives matter	Queries that directly relate to the Black Lives Matter movement	BLM cause, BLM movement, protest updates, BLM news
Protests	Queries that relate to BLM protests	BLM destruction, BLM peaceful protests, BLM protest videos, protest in America, protests going on right now, protests in [city/state], protests in US, Chicago protests, Portland protests
Protesters' information	Queries that relate to protesters' rights and information	What are my rights as a protester, civil rights, civil disobedience, where can and can't I protest, protester arrest, protester information
Police-related	Queries that relate to the police	Police, police brutality, police protests, anti-police, anti-police protests, militarize police, government surveillance
Race issues	Queries that relate to race issues	Race, racism, systematic racism, racial injustice
Coronavirus	Queries related to COVID-19	Second stimulus update, COVID, COVID-19 deaths nationally, deaths in my state, availability of COVID-19 testing
Government or Election related	Queries about the administration or presidential elections	why won't the government do anything, Washington, Trump/Republicans, Trump, Joe Biden

A2. Queries used to scrape Google search results

Support queries
1. ACAB
2. are police that have killed mercilessly being punished?
3. can police or local leaders tell us to disperse?
4. civil disobedience
5. no justice no peace
6. systematic racism
7. police defund bill
8. donation links BLM
9. equal rights for Black lives
10. how many BLM protests have been peaceful
11. militarize police government surveillance
12. police brutality during protests
13. police reform
14. racial injustice protests
15. what are my rights as a protester?
16. where can and can't I protest?
17. racial equality

Oppose queries
1. all American lives matter
2. All lives matter
3. Black violence
4. Blue lives matter
5. Cities burning
6. current riots
7. Dangerous from protesters
8. destruction of private properties
9. looting
10. looting and vandalism in [city]
11. police funding
12. police injuries
13. protesters and crimes
14. protesters hurting others
15. Antifa
16. tearing down of historical statues
17. violent protesters damaging properties

Validation of Twitter political bias scores with other sources



FB Score: Facebook audience-based scores (Bakshy et al., 2015);
AllSides Score & Community Score: Domains rating by AllSides website