The Case for Voter-Centered Audits of Search Engines During Political Elections

Eni Mustafaraj
Department of Computer Science
Wellesley College
emustafaraj@wellesley.edu

Emma Lurie
School of Information
University of California, Berkeley
emma_lurie@berkeley.edu

Claire Devine
Undergraduate Division
Wellesley College
cdevine@wellesley.edu

ABSTRACT

Search engines, by ranking a few links ahead of million others based on opaque rules, open themselves up to criticism of bias. Previous research has focused on measuring political bias of search engine algorithms to detect possible search engine manipulation effects on voters or unbalanced ideological representation in search results. Insofar that these concerns are related to the principle of fairness, this notion of fairness can be seen as explicitly oriented toward election candidates or political processes and only implicitly oriented toward the public at large. Thus, we ask the following research question: how should an auditing framework that is explicitly centered on the principle of ensuring and maximizing fairness for the public (i.e., voters) operate? To answer this question, we qualitatively explore four datasets about elections and politics in the United States: 1) a survey of eligible U.S. voters about their information needs ahead of the 2018 U.S. elections, 2) a dataset of biased political phrases used in a large-scale Google audit ahead of the 2018 U.S. election, 3) Google's "related searches" phrases for two groups of political candidates in the 2018 U.S. election (one group is composed entirely of women), and 4) autocomplete suggestions and result pages for a set of searches on the day of a statewide election in the U.S. state of Virginia in 2019. We find that voters have much broader information needs than the search engine audit literature has accounted for in the past, and that relying on political science theories of voter modeling provides a good starting point for informing the design of voter-centered audits.

CCS CONCEPTS

• **Information systems → Web search engines**.

KEYWORDS

algorithm audits, search engines, Google, voters, elections, bias

ACM Reference Format:
Eni Mustafaraj, Emma Lurie, and Claire Devine. 2020. The Case for Voter-Centered Audits of Search Engines During Political Elections. In *Conference on Fairness, Accountability, and Transparency (FAT* '20), January 27–30, 2020, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3351095.3372835

FAT '20, January 27–30, 2020, Barcelona, Spain*

ACM ISBN 978-1-4503-6936-7/20/02…$15.00
https://doi.org/10.1145/3351095.3372835

1 INTRODUCTION

A week after the 2016 U.S. presidential election, a news story[1] about a problematic Google search result page made the media rounds. The journalist had searched for the query "final vote count 2016", and Google's top search result had claimed that "Trump won both [the] popular [vote] and electoral college." (President Trump lost the popular vote by 2.87 million votes.[2])

This top search result was published by a conspiracy blog, one of the dubious sources that contributed to the creation and spread of a conspiracy theory (repeated by President Trump) that Trump would have won the popular vote if not for voting irregularities (unauthorized immigrant voters, voter fraud, etc.).[3]

How did a false news story from a conspiracy blog end up as the top news story on the Google search results page? Google's response, through a spokesperson,[4] was non-revealing:

> The goal of Search is to provide the most relevant and useful results for our users. In this case we clearly didn't get it right, but we are continually working to improve our algorithms.

Other questions that can be asked in this context are: how many users searched Google for this topic ("final vote count 2016")? How many users saw the conspiracy blog story as the top result? What fraction of users clicked to read it? These questions, aimed at transparency, are important when trying to understand how disinformation spreads and affects users, in order to mitigate its potential harm. For example, Twitter notified 1.4 million users in the United States, that they were exposed to content generated by Russia's Internet Research Agency during the 2016 U.S. Election.[5] Although journalists have documented several examples of disinformation[6] or other harmful content[7] being displayed at the top of Google Search results, Google's response has been to change the algorithms to fix the particular problem, without providing transparency to stakeholders.

This lack of transparency about the *how* of such scenarios is often defended as a protective measure: malicious third-party actors could benefit from efforts of transparency to increase the success

[1] https://www.mediaite.com/uncategorized/now-even-google-search-aiding-in-scourge-of-fake-inaccurate-news-about-election-2016/
[2] https://en.wikipedia.org/wiki/2016_United_States_presidential_election
[3] New York Times: Trump Repeats Lie about Popular Vote in Meeting with Lawmakers (January 23, 2017).
[4] https://www.theverge.com/2016/11/14/13622566/google-search-fake-news-election-results-algorithm
[5] https://www.reuters.com/article/us-twitter-russia/twitter-notifies-more-users-exposed-to-russian-propaganda-idUS.KBN1FK388
[6] https://theoutline.com/post/1192/google-s-featured-snippets-are-worse-than-fake-news
[7] https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook

rate of their attacks.[8] However, Google's secrecy about the extent of exposure to such highly ranked disinformation is potentially harmful to the public. In at least one highly-publicized case, one individual's path to a devastating hate crime—that of Dylan Roof, who killed nine black people in their own church—started with the search for "black on white crime" on Google.[9]

There is a difference between the two search phrases we have discussed so far: "final vote count 2016" and "black on white crime". The first one is a seemingly neutral query that anyone can perform innocently, out of curiosity. The second query is problematic from the start as it was pushed on the Internet discourse by groups with a white supremacist agenda.[10] The researchers Michael Golebiewski and danah boyd coined the concept "data void" [17] to refer to phrases like this one, which, when searched, lead to either low quality or problematic content. As they explain, once "black on white" was publicized by Roof's crime, high quality websites created content that filled the "data void".

One could argue that these are isolated cases, that the majority of queries are not problematic and will not lead to problematic content. However, Google's general lack of transparency has opened the door for continuous political attacks. Most recently, President Trump accused Google of manipulating the search results to favor his 2016 political opponent, Hillary Clinton,[11] basing his claim on a white paper of a small-scale Google audit, which had hypothesized the possibility of Google swaying the elections.

If Google is unable to provide transparency, researchers can try to use audits as a tool for increasing literacy around the complexity of generating search results. The public should be able to search for anything, problematic content as well, but it should be Google's responsibility to not serve harmful content, at the very least, at the top of the search result page. Google was able to fix the issues of consumer-harming web spam that plagued online shopping results in the early 2000's [33]. We should demand the same kind of commitment to high-quality results for political content, too.

This is particularly important, because research has found that the public believes that Google shows trustworthy information at the top of its search results [34]. However, if the 2016 U.S. presidential election was any indication, efforts to propagate disinformation in novel ways will only increase in frequency and sophistication [28]. Continuous and large-scale audits can serve to raise awareness about vulnerabilities in the information ecosystem, for which search engines are a central gateway.

1.1 Auditing for Bias

Social scientists have been worried about the power of search engines and their potential bias since they came into prominence in the 2000's [13, 15, 20, 48]. Thus, auditing search engines in general, and especially their role in political elections, is not a new research problem. However, the underlying assumptions that have motivated such research over the past ten years have been different and led to specific approaches. Concretely, we can distinguish three different kinds of efforts to detect political bias in search platforms:

1. **Third-party manipulation**. In the early 2000's, Google was susceptible to forms of political activism that came to be known as "Google bombing".[12] According to [30], Democratic party activists used this technique to promote in the Google search results negative stories about Republican candidates in the 2006 U.S. Congressional elections. However, such techniques didn't succeed in 2008, due to Google's changes in its algorithms to promote official sources when searching for candidate names. A repetition of this study (with more search engines) in 2016 [31] confirmed the previous finding of official websites topping the rankings.
2. **Ranking Bias**. Experimental work by Epstein and Robertson in [14] tested the following hypothesis: were a search engine like Google to subtly manipulate the ranking of stories about political candidates, this would be sufficient to affect the outcome of an election. They named this the "search engine manipulation effect" (SEME). This hypothesis is connected to the popular theory of "filter bubbles" by Eli Pariser [35], according to which, different search engine users might be exposed to different search results for the same query. Various studies to test this theory (using political search terms, although not during election periods) have found little evidence for it [21, 37, 39].
3. **Ecosystem bias**. Recent audit studies have recognized that focusing on the the producers of content or on the ranking algorithms individually is not sufficient to appropriately characterize the complex nature of interactions during search, which involve the users who initiate the search, the content providers, and the ranking algorithms. Furthermore, the search page itself has evolved over the past years, providing more outlets for showcasing content without having to click on any links on the search page. This ecosystem approach to bias was articulated first in [23] in the context of Twitter search, but can also be found in [11, 19].

As we can notice from this chronological literature survey, there has been a progressive move toward expanding the focus of where bias (and as result, manipulation) is located. For example, the work by Kulshrestha et al. 2017 [23] introduced the concept of the interaction among various kinds of biases (input bias and ranking bias) in order to produce a more pronounced output bias. Meanwhile, the work by Hu et al. 2019 [19] changed how audits are performed, by adopting "biased searches" as starting points for the audit. This allowed them to notice a similar effect to that shown by Kulshrestha et al. on Twitter, namely, that when using biased search phrases, the text snippets shown by Google are more biased than the web pages from which they are extracted.

1.2 Considering the Voters

This progress on auditing search engines in the context of political elections, especially recognizing the many sources of bias and their interaction, is important, but insufficient towards the goal of developing voter-centered audits. For one, most of the mentioned studies utilized queries chosen by researchers themselves, without any

[8]https://www.theguardian.com/commentisfree/2016/nov/13/good-luck-in-making-google-reveal-its-algorithm
[9]https://www.npr.org/sections/thetwo-way/2017/01/10/508363607/what-happened-when-dylann-roof-asked-google-for-information-about-race
[10]https://www.gq.com/story/dylann-roof-making-of-an-american-terrorist
[11]https://www.nytimes.com/2019/08/19/us/politics/google-votes-election-trump.html
[12]https://en.wikipedia.org/wiki/Google_bombing#Political_activism

input from voters or insights from political scientists' knowledge of voters. Additionally, queries are focused on politicians (occasionally, rightly so [16]), while elections are multi-faceted; what else besides the candidates affects participation and decision making in elections? Meanwhile, what role does the interface of the search engine (e.g., the autocomplete suggestions feature) play in shaping voters' information seeking behavior? Finally, do the audits consider if searches are leading to high quality information from reliable sources? This last question is especially important, because efforts to manipulate elections might not come in the form of political bias, which is what most audits have been focusing on. Efforts to suppress voters' participation in elections can target information such as: when does the early voting start? Where are the polling locations situated? When is the voter registration deadline? Ensuring that all voters get reliable information on such questions is important, because such information is location dependent.[13]

In this paper, we explore four different datasets that capture various aspects of searches about politics and elections, with the **explicit goal of identifying important features for the design of search engine audits that are centered on fairness to voters and their information needs.** Our takeaways, explained in more detail in the rest of the paper, are the following:

- **Information Cues:** We need to utilize the vast political science theory on voter modeling to choose queries for audits. For example, we found empirical evidence that many voters behave as predicted by the theory of "information cues" by Arthur Lupia [25–27]. This theory suggests that voters prefer to take shortcuts to get informed about elections. Examples are searches for endorsements or political alignment. Women candidates are more likely to be the subject of such searches.
- **Biased Searches:** Recent audits like Hu et al. [19] are right about including biased searches in their seed dataset of phrases. Our survey with voters found evidence that voters do indeed perform biased searches. However, comparing the phrases that Hu et al. extracted from politician speeches with those used by voters showed almost no overlap. Biased searches need to come from voters.
- **Beyond Candidate Names:** While candidates are central to the information seeking process, queries about them are more complex than just their names. Voters first need to find out who the candidates are, e.g., "Colorado candidates", and then will ask for more specific information, such as, "what does Beto stand for" or "Gillum - DeSantis polling". Moreover, questions about the logistics of voting and the nature of elections are common too.
- **Unreliable Localization:** Voters might prefer short and ambiguous queries, e.g. "election" or "voting near me", and rely on the search engine to either provide autocomplete suggestions or display the right information. Google's algorithms appear to be sensitive to real-world events such as elections, but this behavior seems difficult to predict.

We hope that these insights will serve to guide the design of future voter-centered audits, which, by taking an ecosystem view, aim to uncover whether voters are receiving reliable information from trustworthy sources, independently of their geographic location or the sophistication of their search phrases.

[13]This is especially true in the United States, where each U.S. state has different elections norms and laws.

2 RELATED RESEARCH

Given our focus on the voters, we start this section with theories that try to explain and predict voter behavior with respect to their information needs. Then, to expand upon the overview of audits in the context of political elections, which we presented in the introduction section, we summarize aspects of the audit methodology that is relevant to our discussion.

2.1 Informed or Uninformed Voters?

In 1960, the landmark study of "The American Voter" [7] changed the way the political science research community approached the understanding of electoral behavior, by borrowing methods from the field of psychology. One of the important insights drawn from this new methodology was the discovery of a stable "party identification" as the principal factor in voting. Further exploration of this study's panel data by Philip Converse, exposed the various belief systems of different groups of voters [9]. Many interpreted this research as suggesting that "voters are dumb," and Converse later tried to distance himself from such an interpretation [8], insisting that he had only made the case for a great (mal)distribution of political information across the public, that is, a majority of the public knows little and is inconsistent in their positions (uninformed or low-information voters); while only a fraction of the electorate has sufficient knowledge to be consistent in their positions over time.

A more sympathetic view of the voters comes from Arthur Lupia, who in earlier studies [25, 26] contended that voters could not be expected to have an encyclopedic knowledge about politics. He suggested that information cues (e.g., who supports a ballot initiative, Ralph Nader or the auto insurance lobby? Or, is the candidate a Democrat or Republican?) are a good way for voters to align themselves with the issues or candidates (see also, [24, 32]). A cue, according to Lupia [27] is "a piece of information that can take the place of other information as the basis of competence at a particular task." Cues are not unique to politics, individuals use them in many situations, for example, brand names (e.g. Nike or Volvo) are a cue for product quality. In the survey data we collected, many queries formulate an explicit information request for party identification, e.g., "what party is Harry Arora", "Gavin Newsom's party", as well as other information cues such as endorsements and polls. We qualitatively code all phrases that request information cues to find out the extent this strategy is used by voters.

Additionally, Lupia and other researchers [47] highlight the role that negative emotions such as fear and anxiety play in information seeking behaviors. We noticed this emotional manifestation in searches from our survey participants ("are the elections rigged" or "will democrats win the senate") and decided to look for more examples of such searches, which we broadly categorize as "biased" searches, because they come from an one-sided perspective, usually in favor or against a certain political party, individual, or issue.

2.2 More on Search Engine Audits

Search engines, as one of the major platforms for satisfying information needs, are a particularly frequent target of algorithm audits.

But how are audits carried out? In [40], Sandvig et al. formally define four different types of audits: 1) code audits, 2) scraping audits, 3) sock puppet audits, and 4) non-invasive user audits. **Code audits** involve researchers examining the code bases that underlie algorithmic systems. To date, there are no such audits for search engines, and most likely, the complexity and scale of their code bases might make that task impossible [6]. Instead, the most common way to perform search engine audits is the **Scraping audit**, during which, automated scripts send batches of queries to a search platform and then analyze the results. The datasets used in this paper are collected via scraping audits.

Recently, **sock puppet audits** have increased in popularity as they add a new dimension: the process imitates users by constructing fake user profiles before making repeated queries. Sandvig et al. express concern that these techniques violate platforms' Terms of Service and present some legal issues. However, the auditing community has argued that the importance of understanding potential discrimination and biases in these algorithms outweighs these risks [10, 22]. As this method evolves, some of the sock puppet audits are now employing real users who have been recruited throw crowdsourcing platforms [37, 39, 45].

The **non-invasive user audits** require users to answer questions about the ways they interacted with the platform. This could include, but is not limited to, search query log collections as well. Sandvig et al. raise several concerns about this technique including: users inability to accurately report their behavior, users not searching enough queries that will reveal platform discrimination ("sampling problems"), and the difficulty to establish causality due to lack of systematic controls over what users do. As a result, such audits are almost non-existent in the search engine audits literature. However, this is a limitation that needs to be overcome through a combination of large-scale representative sampling (as the one used in political science surveys) with advances in browser plugin technologies such as the ones used in [29] or [39].

3 DATA AND METHODS

In this paper, we make use of four different datasets. An overview of the datasets is provided in Table 3. Each subsection provides information about how we collected and processed the datasets. More details about the survey with AMT workers, the audit method, and the Virginia election are provided online.[14]

3.1 Dataset 1: Voter Searches

Knowing what voters search ahead of elections and what results they are shown by a search engine can help us ascertain that search engines are not promoting disinformation harmful to the public or the candidates participating in the electoral process. One way to do that would be to recruit participants that are willing to install a plugin on their browser (a technique used by [29, 39]), which will allow us to collect their interactions with the search engine.

However, implementing plugin based studies is expensive, introduces concerns about user privacy, and presents new technical challenges in having to discern which queries are election related. Therefore, we created a survey rather than a plugin to collect election related queries.

[14] http://cs.wellesley.edu/~credlab/fat2020/

We recruited 560 U.S. citizens through Amazon Mechanical Turk (AMT) and asked them to share searches that they have performed or will perform on the days ahead of the election (Oct 16-Nov 2, 2018). To attract a diverse group of participants, we didn't set restrictions for participation, except for age (over 18) and citizenship (United States). This led to a large number of incomplete or irrelevant submissions that we manually discarded. At the end, we had responses from 392 eligible U.S. voters, who supplied between 3 to 8 search phrases, together with demographic information.[15]

The respondents generated a total of 2,526 search phrases, with 84% of queries only appearing once in the dataset. The average length of a search phrase was 3.2 words (standard deviation = 1.82 words, and median = 3 words). This finding mirrors what studies of search logs across many datasets have found: an average query length of 2-4 words [1]. That is, despite the fact that the queries in this dataset were not extracted from search logs, they still conform to the usual form of search queries. We perform qualitative coding of these search phrases to detect themes, identify biased searches, as well as requests for information cues.

3.2 Dataset 2: Partisan Queries

Hu et al. 2019 [19] performed an audit of Google search snippets during Oct 13-30, 2018, ahead of the 2018 U.S. Congressional Elections. This is the largest audit about elections ever reported in the literature. The authors started with a list of 3,520 politician names and 1,050 left and right leaning terms and phrases. Then, they extracted Google autocomplete suggested phrases for all of them and collected the resulting 88,745 SERPs (search engine result pages). Differently from previous audits in the literature, which mostly rely on candidate names and phrases selected by researchers, this time the politically-biased searches came from politicians. Concretely, the authors created a lexicon of unigrams and bigrams that represent partisan cues. These n-grams came from speeches of 20 Democrats and 19 Republican politicians, scraped from the website `votesmart.org`, and spanning the time period January 2008 - August 2018. All n-grams that occurred more than 50 times were assigned a partisan bias score, which ranged from -1 (ngram used only by Democrats) to +1 (ngram used only by Republicans). Then, from this lexicon, the authors chose all phrases with an absolute score greater than 0.5, and manually selected meaningful phrases to search. The dataset that they shared with us contains 495 such phrases and their bias scores. For example, "gun lobby" has a score of -0.85 (mostly used by democrats), while "gun rights" has a score of 0.67 (mostly used by republicans). In our analysis, we search if any of these phrases is contained in the searches formulated by the AMT respondents, in order to discover if such phrases are representative of what voters themselves could search ahead of an election.

3.3 Dataset 3: RS-Candidates

For Google Search users, there are two opportunities to learn what other users are searching: the autocomplete suggestion feature and the related searches at the bottom of a SERP. Autocomplete searches are sensitive to "external shocks" [38] (events that happen in the real world) and are updated more frequently than related searches.

[15] Refer to materials on our webpage: http://cs.wellesley.edu/~credlab/fat2020/

Table 1: An overview of the four datasets used throughout this paper. Three datasets were created by this paper's authors and one was received by the authors of Hu et al. 2019 [19].

	Name	Period	Description	Creator
1.	Voter Searches	Oct - Nov 2018	A list of 2,500 search phrases collected through a survey of AMTs.	Authors
2.	Partisan Queries	Jan 2008 - Aug 2018	A list of 495 phrases extracted from politicians' speeches of democrats and republicans with a partisan bias score.	Hu et al. 2019
3.	RS-Candidates	Oct - Nov 2018	Google's related searches (RS) in SERPs for two groups of candidates.	Authors
4.	Virginia Election	June 11, 2019	Autocomplete phrases and SERPs for 40 queries searched simultaneously in Virginia and Massachusetts.	Authors

One way to think about these two groups of searches is that the former correspond to "trending searches", while the latter to "most popular searches" about a seed query.

After collecting SERPs for two groups of candidates for the 2018 U.S. midterm elections, we extracted the related searches (RS) from each page. The groups of candidates were as follows:

(1) 185 Women candidates [185Women] running for congress or governor. Their names were gathered from a *Washington Post* article that was tracking the political fortunes of women candidates.[16]
(2) 215 Senate challengers [215Challengers], running in the primaries and the 2018 Senate election for 35 U..S. states. The names of candidates were collected from Ballotpedia in July 2018, by visiting the electoral information webpage of each US state holding a 2018 election.

An example of related searches for the candidate Alexandria Ocasio-Cortez is shown in Figure 1. For our analysis, we are interested in the words (or phrases) next to the candidate name, in this case, "husband", "married", "age", etc. To have a more complete dataset, we collected SERPs for both groups of candidates on 6 different days on the period Oct 17, 2018 - Nov 3, 2018, and compiled all unique phrases in related searches for each candidate. Then, we automatically removed the names of the candidates to focus on the additional words and phrases in the queries. We found 1,944 unique phrases for [185Women] and 1,974 unique phrases for [215Challengers]. These phrases were manually inspected to correct some of the inevitable errors of the automatic phrase extraction.

Figure 1: The related searches for Alexandria Ocasio-Cortez (a woman candidate running for the U.S. Congress), screenshot taken on Nov 3, 2018. We automatically remove the candidate name from the related searches and aggregate the remaining words or phrases. In this case, these words are, husband, married, age, etc.

[16] https://www.washingtonpost.com/graphics/2018/politics/women-congress-governor/

3.4 Dataset 4: Virginia Election

When we analyzed the phrases from the *Voter Searches* dataset, we noticed that many of them were short and ambiguous. For example, respondents wrote phrases like: *voting* or *elections*, as well as "candidates near me". We hypothesized that one reason that voters might formulate such queries is that they rely on the search engine to correctly interpret their intentions, despite their searches being short and ambiguous. Google users are familiar with searches like "weather" or "pizza near me" that correctly leverage the user's location to personalize the search results [18]. Another reason for supplying short queries to the search engine might be that users are accustomed to taking advantage of Google's autocomplete feature to suggest longer and more appropriate queries. In order to test such hypotheses, we conducted a pilot study the day of the primary election for the State Legislature in the U.S. state of Virginia, June 11, 2019. We selected 40 query phrases from the *Voters Searches* dataset, mostly short queries such as: *voting*, *election day*, *democrats*, as well as syntactic variations of semantically similar queries such as *where can i vote*, *where do i vote*, or *where to vote*.[17] We used two computers, one in Virginia and one in Massachusetts to automatically collect the SERPs for these search queries from Google at the same exact time and through the same automated procedure.[18] Additionally, we collected the autocomplete suggestions for all the queries (10 for each) in both locations, using the open-source tool[19] introduced in [38]. Thus, this dataset is composed of 80 SERPs (40 for each location), and 800 autocomplete search phrases (400 for each location).

3.5 Qualitative Coding

For the *Voter Searches* dataset, we performed multiple rounds of qualitative coding. We started with a thematic analysis of the data [4], in order to inductively draw themes from the queries. Then, we performed deductive coding for the presence of phrases that elicit information cues (yes, no), presence of bias (yes, no), and expression of bias (semantic, pragmatic).

For the thematic analysis, two coders independently looked for themes in a subset of 10% of the phrases, until they reached a saturation point. The coders then met to discuss themes and subthemes and create the code book. The rest of the data was thematically coded by a single coder. For the deductive coding of information cues, bias presence and bias expression, two independent coders

[17] The list of all queries can be found online: http://cs.wellesley.edu/~credlab/fat2020/.
[18] Our procedure is described here: http://cs.wellesley.edu/~credlab/fat2020/.
[19] https://github.com/gitronald/suggests

and the research team defined strategies and rules for applying the codes and then the coders proceeded independently. After each round, they had the opportunity to check their differences and decide to change or retain their label. We report the inter-rater reliability (IRR) score, calculated as the Cohen's kappa in the Results section.

Operationalizing Bias: Our definition of what is biased was deliberately broad: any explicit or implicit expression of preference toward or against a person, group, issue, or event. We recognize that the discussion of bias in literature is often much more narrow and usually with a negative connotation. However, to the extent that bias seems to be correlated with information cues, which, when accurate, can help voters make good decisions, we think that our broad definition is warranted. Particularly in psychology, many biases, e.g., optimism bias [41], are shown to have positive effects on our lives. Inspired by this interpretation, we don't regard the presence of bias in search queries (the phrases provided by voters) negatively.

Once we had identified the subset of biased phrases, we proceeded with a second round of coding, this time to identify the expression of the bias using a binary categorization: semantically and pragmatically biased queries. A query was coded as *semantically biased*, if it included language that denotes bias independently of its context. For example, the phrases "the best candidate" and "will Beto win" are semantically biased because of the individual words "best" and "win." Alternatively, a query was coded as *pragmatically biased*, if it contains bias as a result of its context within a broader narrative. Such examples would include "diane feinstein's age" or "blue wave." None of the words in these phrases indicates bias, but in the context of the election, the phrases can be interpreted as biased. Concretely, there were calls for California Senator Diane Feinstein to not run again for a senate seat, because she was deemed too old, leading to accusations of ageism. Meanwhile, "blue wave" referred to a theory in which the Democrats (whose color is blue) were going to regain the power in Congress, leading to much excitement and mobilization for Democratic Party voters.

Our terminology (semantic/pragmatic) was inspired by the "relevance theory of meaning" within linguistics and philosophy of language scholarship. Developed in the 1980s by Dan Sperber and Deirdre Wilson [42], relevance theory distinguishes between the semantic and pragmatic meaning of a linguistic utterance in order to describe how people interpret context-independent linguistic cues based on speaker intention and contextual situation.

Operationalizing Information Cues. Drawing on Arthur Lupia's discussion of information cues in his book [27], we identified queries which likely contained shortcuts to information that voters may not have on recall. As a reminder, Lupia explains that "a cue is a piece of information that can take the place of other information as the basis of competence at a particular task." Our dataset doesn't contain the cues themselves, instead it contains the query phrases that voters formulate in order to access such cues. For example, when they search about "Gavin Newsom's party", they are looking for this candidate's partisan alignment that serves as a strong information cue. Other queries that serve to access information cues are endorsements, polls, rankings, or position alignments with hot-button issues.

3.6 Limitations of the Collected Datasets

All our datasets can be considered as "small" datasets compared to the ones used in other auditing studies. This was by design, so that we could inspect the results manually, and only small datasets are amenable to such analysis. Our goal is not to perform an audit, but to find out how to design audits that are centered on voters. We believe that the exploration of these small datasets provides important insights that can be tested via large-scale and partially automated audits. Nevertheless, it's worth pointing out the limitations of the datasets, in order to mitigate such limitations in future studies.

Sampling Bias: The survey with AMT respondents is not representative of the voter population in the United States. However, recent research has indicated that AMT is better than or equal in terms of diversity to other survey participant pools [3].

Ecological Validity: We didn't receive actual search logs from the respondents, thus, it's fair to question how valid the query phrases are. As part of the survey, we asked participants how likely it was that they have performed or will perform these searches and 79.3% responded with extremely likely or somewhat likely. The average length of phrases was 3.2 words, within the range described in literature [1]. It is possible that when searching with a search engine, voters will formulate different queries, but, we believe that this doesn't invalidate that the queries in our dataset are possible, too. Given that this study only uses the data to inform design for future audits, the incompleteness of the dataset might be acceptable.

Algorithmic Exclusion: Two datasets, *Virginia Election* and *RS-Candidates*, are based on data that Google's algorithms for generating related searches and autocomplete suggestions provide. It is the nature of these algorithms to exclude certain terms and phrases that are deemed harmful.[20] Relying on accounts of researchers who have accessed unrestricted search logs [44], we know that users perform many searches that are not made public in any way. While the exclusion of such queries limits what we can learn from the datasets, it is also a reminder of why we cannot perform audits relying only on autocomplete searches, but instead find ways to collect real queries from users, given that access to search engine logs [49] is off limits for most researchers.

4 RESULTS

One central motivation for this study is to gather information for designing voter-centered audits of search engines during election periods. The analysis of the four datasets we have assembled provides us with the opportunity to do so.

4.1 Information Cues

Our analysis of information cues, phrases that serve as shortcuts to evaluate candidates, explores two datasets: *Voter Searches* and *RS-Candidates*. In *Voter Searches*, we manually coded all phrases as eliciting information cues or not. The first rater coded 32% (673 phrases) and the second rater coded 37% (789 phrases) as eliciting such cues. Their agreement as calculated by Cohen's kappa was 0.7 (substantial agreement). By matching the phrases back to the respondents, we found that 264 participants (67%) formulated at least one query that elicits information cues about candidates or

[20] https://www.blog.google/products/search/how-google-autocomplete-works-search/

Table 2: Comparing the top most related searches for two datasets: [185Women] and [215Challengers], collected during the 2018 U.S. Congressional Elections. Searches known as *information cues*, such as *polls* or *endorsement*, are more often directed at women candidates.

[185Women] Dataset			**[215Challengers] Dataset**	
related searches	%		related searches	%
polls	**55%**		senate	41%
facebook	45%		for senate	39%
wikipedia	41%		twitter	23%
bio	38%		facebook	21%
for congress	38%		wikipedia	20%
age	35%		**polls**	**19%**
twitter	35%		bio	15%
endorsements	**30%**		us senate	15%
husband	26%		age	15%
congress	24%		net worth	15%

issues, with 97 participants (25%) formulating three or more such queries. Thus, 2/3 of participants engaged in this kind of information gathering strategy, as Lupia's theory on voters suggests. This is important on two counts: a) it indicates that we can rely on political science theory for voter modeling to extract insights to inform audits; b) it raises the issue of whether the results that are shown in response to such queries are reliable, given that voters are using them to make decisions about who or what to vote for.

While *Voter Searches* coverage is restricted to the convenience sample that we recruited through Amazon Mechanical Turk (n=392), related searches for candidates on Google Search might correspond to a larger sample, accumulated over weeks and months.[21] Thus, it is informative to analyze what Google users most frequently search about candidates. Table 2 summarizes the top phrases that we extracted for the two sets of candidates, [185Women] and [215Challengers]. The value in percentage indicates the proportion of candidates for which the query was found at least once in their related searches. For example, searches about polls were found in 55% of [185Women] and in 19% of [215Challengers]. From the perspective of the information cues theory, such results are interesting because they not only indicate that Google users are looking for them (for example, polls and endorsements, shown in bold), but that such requests are much more likely for women candidates. While many of the phrases in Table 2 can be regarded as navigational queries (based on Broder's taxonomy, [5]), some of the information queries formulated for women candidates (age or husband) are a good indicator of the personal scrutiny that women running for office face, another reason for considering the presence of bias by voters in search phrases.

4.2 Bias in Search Phrases

Our analysis of bias was focused on two datasets, *Voter Searches* and *Partisan Queries*. Two independent raters labeled phrases in *Voter Searches* as "biased" or not. In the first round of labeling they achieved an inter-rater agreement (Cohen's kappa) of 0.68 (substantial agreement) and in the second round of reconciliation, the agreement rose to 0.88 (almost perfect agreement). At the end of these two rounds, the two raters agreed that 657 phrases, or 31% of the whole dataset, is composed of biased search phrases. Matching these phrases to the participants in the AMT survey, we found that 69% (269 out of 392) of the participants have formulated at least one biased search. Meanwhile, 26% (104 out of 392) of participants formulated three or more biased searches. These results indicate that formulating biased queries is a common practice for a majority of voters and that a sizeable portion of them will frequently perform such searches.

For the set of 657 biased phrases, the raters performed another round of coding to establish the expression of bias (as semantic or pragmatic). The two raters labeled 59% and 62% (respectively) of phrases as semantically biased, with the rest labeled as pragmatic. Their agreement according to the Kappa score was 0.77 (substantial agreement). The good news is that the majority of the biased searches express this bias semantically, which we defined as being located in one of the query words themselves. These are words that are found in sentiment analysis and opinion mining lexicons, as having a positive or negative valence, e.g, [2], or through NLP approaches [36]. This will make it easier (for large-scale surveys) to identify biased searches, in order to scrutinize the search results shown for them. However, a substantial portion of bias is expressed pragmatically and this will complicate automated analysis.

Although we didn't explicitly set out to find "data voids", while performing the labelling into semantic vs pragmatic bias, we noticed that several of the phrases fit the description of searches that have the potential to lead to "data voids", SERPs with low-quality content from adversarial information providers. The typology of data voids proposed by Golebiewski & boyd [17], would classify some of them as data voids that can be weaponized after breaking news events, e.g, "maga bomber" or "25th ammendment", and others that may be weaponized due to ongoing discriminatory events in our society, e.g., "voter purge" or "immigration based crime." Thus, an advantage of collecting such queries from users is to understand the extent to which certain weaponized phrases succeed to spread in the public.

In the past, most researchers have used neutral phrases as queries for the search engine audits. Hu et al. 2019 [19] intentionally used biased phrases and their autocomplete suggestions to perform their audit. In light of our above-mentioned findings, such a decision is the right one, given that voters will either occasionally or frequently perform biased searches. However, it's worth examining more closely the details of how bias is expressed. For that, we compared the phrases in *Voter Searches* with those in *Partisan Queries*. We found an exact overlap for only 7 phrases. These phrases are (occurrence given in parentheses): early voting (13), gun control (11), illegal immigration (4), voter turnout (3), voter suppression (3), russia investigation (1), election system (1). That is, 7 phrases or 1.4% of biased phrases in *Partisan Queries* (extracted from politician speeches) were found 36 times (1.4%) in *Voter Searches*. This very small overlap is concerning, because it indicates that we cannot rely on the speech of politicians or other political elites as a source of knowledge. Additionally, one reason that Hu et al. 2019's data might be showing this low overlap is that it draws from politician speeches over an extended period of ten years, while the public

[21] However, we have no way of knowing this with any certainty, since Google doesn't divulge absolute numbers about search volume.

Table 3: The 10 major themes that we identified coding the data. Many search phrases are coded with multiple themes, since they don't fit exclusively in one theme only. As a result, the relative frequencies don't add up to 100%.

Theme	Examples of search phrases	Rel. Freq.
Ballot	ballot initiatives in 2018, proposition 110, MA question 1, Oregon measure 103	6.37%
Candidates	Black democratic candidates Florida, is there a female running in Georgia, policies of Alaskan candidates	41.57%
Election	election 2018 reviews, Senate seats up for reelection, Ohio election polls, voter suppression	23.79%
Issue	immigration based crime, georgia hope scholarship Abrams, gillibrand gun control, arizona health care	14.05%
News source	AP midterm facts, ballotpedia, 538 senate forecast, recent election news	3.44%
Office	NC judicial candidates, Virginia Beach mayor, who my senate candidates are, best candidate for governor	12.23%
Party	how can Dems take the house, florida democratic party, Travis county GOP, Republican donation limits	12.51%
People	Gillum vs DeSantis polling, Ben Jealous policies, bernie sanders endorsements, Claire McCaskill	15.52%
State	overview of stances of arizona politicians, california voter guide, Colorado poll, who's winning florida	23.56%
Voting	where can I vote, when do polls end, register to vote Ohio, absentee ballot, voting locations central Michigan	10.65%

might be more tuned to phrases that are currently in the news. Thus, gathering search phrases directly from the public might be a necessary condition for voter-centered audits.

4.3 Beyond Candidate Names

Most audits in the literature have relied on candidate names. While it is true that voters are interested in candidates, their query phrases are not simply candidate names. We found only 391 query phrases that explicitly mention candidates. Out of these, 50% mentioned candidates in context, such as, "Bredeson stance on healthcare", or "elizabeth warren ancestry". Of the other half, 39% referred to complete names, 9% to last names only (e.g., "Gillum") and 2% to first names only (e.g., "Beto"). Most importantly though, initially, only 135 respondents explicitly mentioned a candidate in their phrases. With our prompt, this number increased to 192 (slightly less than half of all participants). Meanwhile, as the results of the thematic coding shown in Table 3 indicate, the largest group of searches is about unnamed candidates, with queries such as "Black democratic candidates Florida" or "policies of Alaskan candidates". This is indicative of another political science theory [9], that most voters are initially uninformed. Thus, the role that the search engine will be playing, by directing voters to information sources in response to such queries, is even more important, given the scale that this might be happening.

The thematic coding in Table 3 contains more insights. The voters are interested not only in a wide range of topics, but also in a variety of information kinds: factual information ("where can I vote"); speculative information ("who's winning Florida"), and so-called problematic queries that might lead to data voids (e.g., "immigration based crime"). Given such variation, to carry out comprehensive voter-centered audits, we will need to design mechanisms to elicit such queries from voters, in order to examine how the search engines treat these distinct epistemic categories (factual, speculative, and problematic) of information needs. Ethnographic research by [46] provides additional evidence for such a need.

4.4 Unreliable Localization

The *Virginia Election* dataset contains SERPs and autocomplete suggestions that were collected on two different locations, Virginia and Massachusetts, on June 11, 2019, the election day in Virginia. Comparing the two sets of 400 autocomplete phrases, we find an overlap of 76.4%. While the overlap is substantial, the fact that 23.6% of phrases are different, indicates that there is a degree of localization happening in both locations. Given that our focus was on the election happening that day, we further analyzed the seeds and suggested phrases that indicated an awareness about local elections. The results for both locations are shown in Table 4 and Table 5. Comparing the results in the two tables shows that there are

Table 4: Out of 40 seed queries, 13 received 23 autocomplete suggestions for the location in Virginia, related to the primary election happening that day in Virginia. (R = Rank)

Seed	Suggestion	R
candidates	candidates in virginia primary 2019	8
election	election day 2019 virginia	1
election day	election day 2019 virginia	1
elections	elections in virginia 2019	1
	elections in virginia	4
	elections in va	6
how to register	how to register to vote in va	1
how to register to vote	how to register to vote in va	2
	how to register to vote in virginia	3
primaries	primaries in virginia 2019	1
	primaries virginia	3
sample ballot	sample ballot virginia	1
	sample ballot fairfax county	4
voter registration	voter registration virginia	3
	voter registration card va	8
voting	voting in virginia	2
	voting in va	6
	voting virginia 2019	7
where do i vote	where do i vote virginia	1
	where do i vote in primaries	2
where to vote	where to vote in virginia	1
	where to vote in virginia primary	2
who is running	who is running for virginia senate	3

Table 5: Out of 40 seed queries, 7 received 13 autocomplete suggestions for the location in Massachusetts. There was no election happening in Massachusetts on the collection date.

Seed	Suggestion	Rank
election day	election day 2019 massachusetts	7
how to register	how to regiser to vote in ma	7
how to register to vote	how to register to vote in ma	2
	how to register to vote in mass	5
primaries	primaries in massachusetts	8
sample ballot	sample ballot brookline ma	6
	sample ballot massachusetts	9
voter registration	voter registration ma	2
	voter registration card ma	5
where do i vote	where do i vote ma	1
	where do i vote boston	2
	where do i vote brookline ma	4

more suggestions for the Virginia location (23 versus 13) and that the suggestions generally appear towards the top of the SERP. This can lead us to hypothesize that to some extent, Google's algorithms are able to pick up the signal about Virginia's elections and update the autocomplete suggestions accordingly.

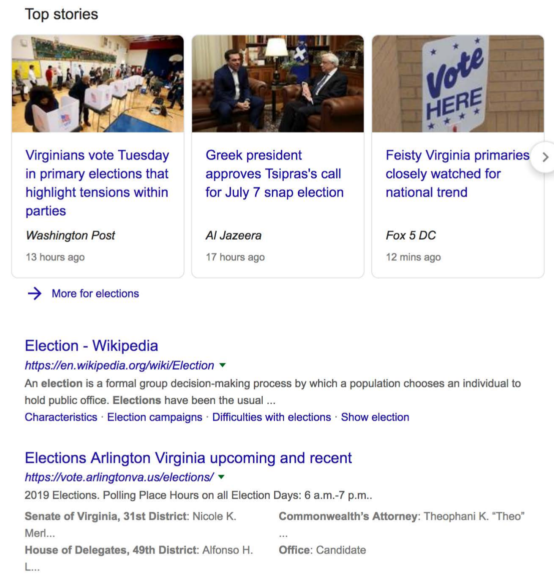
However, most of the seed queries didn't contain any localized suggestions (27 for VA and 33 for MA) and it is difficult to predict which queries will display locally relevant results. This is especially concerning for queries that are semantically similar, but display minor syntactic differences. For example, the phrases "who is running" and "who's running" differ only in the contraction of the verb, but all their suggestions but one are different. Similarly, suggestions for "where can I vote", "where do i vote", and "where to vote" don't overlap much. This random variation makes it challenging to consider autocomplete suggestions as a reliable source for directing searchers toward valuable suggestions during an election period.

On Virginia's election day, we also collected SERPs in both locations. Are SERPs personalized to show results based on location for simple queries such as "elections"? As the screenshot in Figure 2 shows, there is evidence for such hypothesis too.

Nevertheless, drawing a clear conclusion from the comparison of the organic links from all 40 SERPs between the two locations was also challenging. For 8 of the queries, the results are identical (e.g., polls, election polls, republicans, who is running, etc.), but for 9 of them, the Jaccard similarity[22] is less than 0.3 (little overlap between links). Some of the search phrases with little overlap include: "voting locations", "elections", "poll results", "sample ballot", "how to vote", etc. For these queries, many local links (local government or local news) are shown. This is a desirable situation, but the unpredictability of which queries lead to locally relevant SERPs and which do not makes this an unreliable feature, similarly to autocomplete suggestions. How consistently are results for important searches localized? Does it depend on one's locality? Does it depend on the sample size of users who engage in such searches

[22] A similarity measure that is calculated as the ratio of the size of the intersection of the two lists with the size of the union of the two lists. It is 0 for two lists with no overlap and 1 for two lists with full overlap.

Figure 2: Screenshoot of Google's SERP for the query "election', taken on June 11, 2019, from a computer in Virginia. Two articles in Top Stories and the second result are about the Virginia election.



from that location? If so, that might be unfair to voters who live in remote areas (away from densely populated cities) or who lack local institutions that will post on the web reliable information. Further research is needed to addresss such questions.

5 DISCUSSION AND IMPLICATIONS

There is surprisingly little research on how voters use search engines in the context of political elections, see [12] for a recent multi-national survey, supported by Google. Concretely, the long-running American National Election Study (ANES),[23] which has been collecting data on voters and elections in the U.S. since 1948, has yet to include questions about the use of search engines by voters, despite asking questions about Twitter (a much less used platform by the population at large).

The analysis of our datasets, despite the stated limitations, provides insights that should inform further research both on how voters utilize search engines for staying informed during elections and on the kind of results search engines provide.

5.1 Bias, Information Cues, and Data Voids

By labeling all search phrases in our dataset as either "biased" or "not biased," we found that 69% of participants formulated biased searches (toward or against)[24] a political party, ideological issue, candidate trait, or entire groups of people. Often, this bias is evident through partisan words (e.g., Democrat or Republican); positive and negative emotions or situations (e.g., win, lose, good, bad); verbs that pass a judgment (e.g., lie, impeach); or verbs that express

[23] https://electionstudies.org/

[24] We didn't label the direction of bias.

support (endorse, approve, etc.). We refer to all these instances as semantically biased phrases, given that the bias is evident in the meaning of the words. In the biased phrases dataset, close to two thirds of phrases are semantically biased. This is a useful finding because when bias is expressed in this way, it is possible to identify it using automated approaches based on natural language processing techniques. The remaining one third of the phrases were labeled as pragmatically biased, indicating that in order to infer the bias, we need contextual information that is not present in the words themselves. For example, a phrase like "kavanaugh hearing" is pragmatically biased because one needs to know that the way senators voted during the hearings of Supreme Court Justice Brett Kavanaugh became an electoral issue in the November 2018 U.S. midterm campaigns.

Our analysis also illustrates the relationship of biased phrases to two important concepts in political and media communication: information cues and data voids. In our dataset, 75% of queries labeled as biased search phrases were also labeled as eliciting information cues. In addition to well-known cues such as the party affiliation of a candidate or their endorsement by a trusted entity, there seems to be a shift toward a new and broader set of cues. Concretely, we noticed many queries asking for the stance of a candidate on what have become ideologically divisive issues, such as abortion, gun control, marijuana, inequality, climate, etc., which are not always aligned with one's political affiliation. Finding reliable information on the web about such issues might be challenging for many voters. Therefore, Google's decision to display an "On the issues" tab as part of the knowledge panel for the presidential candidates in 2016 and some senators in 2018, may be a positive step toward solving this issue. However, some researchers have criticized the approach, because it relies on biased news sources, and it might not be available for all candidates running for office.[25]

Another type of biased searches that we identified (as part of the semantic/pragmatic labeling of bias) are rumors or conspiracy theories that have the potential to lead to so-called "data voids," situations in which the search engine only shows results from low-quality information sources, because such rumors are not covered from reliable news sources. As discussed by Golebiewski and boyd [17], users are often nudged to search for certain phrases (e.g., "caravan," "immigration based crime," "voting fraud," "voter purge," etc.), by trails left on other media (Twitter, talk show radio, YouTube) from different political actors with varying agendas. Since such rumors, often related to current events, go viral unexpectedly (e.g. "maga bomber" or "anti trumper shoots up synagogue"), identifying them when auditing search results is a challenging task. However, to the extent that they might influence elections, it is a topic that we believe needs further attention by the research community.

5.2 Implications for Voter-Centered Audits

Here are some important takeaways to consider when designing voter-centered audits in the context of political elections:

Acknowledge Bias: Voters perform biased searches, but their expression of bias doesn't match that of established politicians. Fortunately, most of their biased searches are semantically expressed and can be discovered automatically through NLP techniques. Meanwhile, pragmatically biased phrases are difficult to recognize and interpret and may occasionally lead to "data voids." Finding ways to gather/discover such queries and audit their results in a timely fashion should be an important area for future work, especially in the context of fighting political disinformation by bad-faith actors.

Candidates: When voters search for specific candidates, they don't simply use the candidates names. Instead, they formulate specific questions that use the names in context. For some groups of candidates, e.g., women, there are common pieces of information being asked, which might reveal greater bias toward them. It's thus worth considering auditing for groups of candidates in case they are target of disinformation that might be visible only when results are compared to other groups of candidates.

Cues: Voters are formulating searches to lead them to information cues. In addition to well-established cues such as partisanship, endorsements, and polls, a new set of cues (stances on specific issues, not clearly aligned with partisanship) are emerging. This need for quick access to such cues raises the issue of the authority and the political interests of sources that are providing the answers on Google. Methods for assessing who is a trustworthy source in such contexts need to be established.

Localization: Virginia is the 12th most populated state in the U.S., with 8.5 million inhabitants. Their election searches might have been easy to pick up by Google's algorithms. But how do these algorithms behave in other parts of the country (or other countries in the world) on the days ahead of the election? Large-scale audits that target diverse geographical areas, following the method in [21] are needed to ensure that voters in these areas have access to reliable information as well.

6 CONCLUSION

Search engines are one of the most used platforms for accessing information about elections [12]. Our exploratory, qualitative analysis of four datasets related to elections in the United States indicated that 2/3 of voters formulate queries to elicit information cues about elections (shortcuts to information that helps them decide). Similarly, voters also perform biased searches as well as problematic ones. It is thus important that future search engine audits go beyond identifying whether their ranking algorithms are biased, but instead, take a broader ecosystem approach. This means that audits should specifically target the quality of information in response to varied queries (tested in different geographical locations), in order to detect and measure possible pollution in the informational ecosystem, which often is the result of deliberate disinformation by bad-faith actors [43], who are engaged in information warfare.

ACKNOWLEDGMENTS

We would like to thank Jennifer Chudy and Cassandra Pattanayak of Wellesley College, for insightful conversations and pointers to the literature in political science, as well as discussions of survey analysis. We are indebted to Ronald E. Robertson for generous feedback on previous drafts of this article and for continuous inspiration with his work. Finally, we are grateful to the members of the Wellesley Cred Lab for their moral support and to funding from the National Science Foundation, under grant IIS 1751087.

[25] https://slate.com/technology/2016/06/how-the-google-issue-guide-on-candidates-is-biased.html

REFERENCES

[1] Avi Arampatzis and Jaap Kamps. 2008. A study of query length. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 811–812.

[2] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, Vol. 10. 2200–2204.

[3] Frank R Bentley, Nediyana Daskalova, and Brooke White. 2017. Comparing the Reliability of Amazon Mechanical Turk and Survey Monkey to Traditional Market Research Surveys. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1092–1099.

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[5] Andrei Broder. 2002. A taxonomy of web search. In *ACM SIGIR Forum*, Vol. 36. ACM, 3–10.

[6] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* 3, 1 (2016). https://doi.org/10.1177/2053951715622512

[7] Angus Campbell, Philip Converse, Warren Miller, and Donald Stokes. 1960. *The American Voter*. University of Chicago Press.

[8] Philip E. Converse. 2000. Assessing the Capacity of Mass Electorates. *Annual Review of Political Science* 3, 1 (2000), 331–353. https://doi.org/10.1146/annurev.polisci.3.1.331

[9] Philip E. Converse. 2006. The nature of belief systems in mass publics (1964). *Critical Review* 18, 1-3 (2006), 1–74. https://doi.org/10.1080/08913810608443650

[10] Nicholas Diakopoulos. 2015. Algorithmic Accountability. *Digital Journalism* 3, 3 (2015), 398–415. https://doi.org/10.1080/21670811.2014.976411

[11] Nicholas Diakopoulos, Daniel Trielli, Jennifer Stark, and Sean Mussenden. 2018. I Vote For: How Search Informs Our Choice of a Candidate. In *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. Springer, 121–133.

[12] William H. Dutton, Bianca Reisdorf, Elizabeth Dubois, and Grant Blank. 2017. Search and Politics: The Uses and Impacts of Search in Britain, France, Germany, Italy, Poland, Spain, and the United States. *Quello Center Working Paper No. 5-1-17* (2017). http://dx.doi.org/10.2139/ssrn.2960697

[13] Niva Elkin-Koren. 2000. Let the crawlers crawl: On virtual gatekeepers and the right to exclude indexing. *U. Dayton L. Rev.* 26 (2000), 179.

[14] Robert Epstein and Ronald E Robertson. 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 (2015), 4512–4521.

[15] Susan Gerhart. 2004. Do Web search engines suppress controversy? *First Monday* 9, 1 (2004). https://doi.org/10.5210/fm.v9i1.1111

[16] Tarleton Gillespie. 2017. Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. *Information, Communication & Society* 20, 1 (2017), 63–80.

[17] Michael Golebiewski and danah boyd. 2018. Data Voids: Where Missing Data Can Easily Be Exploited. (2018), 1–7. https://datasociety.net/wp-content/uploads/2018/05/Data_Society_Data_Voids_Final_3.pdf

[18] Aniko Hannak, Piotr Sapiezynski, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. 2013. Measuring Personalization of Web Search. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 527–538.

[19] Desheng Hu, Shan Jiang, Ronald E. Robertson, and Christo Wilson. 2019. Auditing the Partisanship of Google Search Snippets. In *Proceedings of the 2019 Web Conference (WWW 2019)*. San Francisco, CA.

[20] Lucas D Introna and Helen Nissenbaum. 2000. Shaping the Web: Why the politics of search engines matters. *The information society* 16, 3 (2000), 169–185.

[21] Chloe Kliman-Silver, Aniko Hannak, David Lazer, Christo Wilson, and Alan Mislove. 2015. Location, Location, Location: The Impact of Geolocation on Web Search Personalization. In *Proceedings of the 2015 Internet Measurement Conference (IMC '15)*. ACM, New York, NY, USA, 121–127. https://doi.org/10.1145/2815675.2815714

[22] Joshua A Kroll, Solon Barocas, Edward W Felten, Joel R Reidenberg, David G Robinson, and Harlan Yu. 2016. Accountable algorithms. *University of Pennsylvania Law Review* 165 (2016), 633.

[23] Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P Gummadi, and Karrie Karahalios. 2017. Quantifying search bias: Investigating sources of bias for political searches in social media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 417–432.

[24] Milton Lodge and Ruth Hamill. 1986. A Partisan Schema for Political Information Processing. *American Political Science Review* 80, 2 (1986), 505–519. https://doi.org/10.2307/1958271

[25] Arthur Lupia. 1992. Busy Voters, Agenda Control, and the Power of Information. *The American Political Science Review* 86, 2 (1992), 390–403. http://www.jstor.org/stable/1964228

[26] Arthur Lupia. 1994. Shortcuts Versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections. *American Political Science Review* 88, 1 (1994), 63–76. https://doi.org/10.2307/2944882

[27] Arthur Lupia. 2016. *Uninformed: Why people know so little about politics and what we can do about it.* Oxford University Press.

[28] Alice Marwick and Rebecca Lewis. 2017. Media manipulation and disinformation online. *New York: Data & Society Research Institute* (2017).

[29] Connor McMahon, Isaac Johnson, and Brent Hecht. 2017. The substantial interdependence of Wikipedia and Google: A case study on the relationship between peer production communities and information technologies. In *Eleventh International AAAI Conference on Web and Social Media*.

[30] P Takis Metaxas and Eni Mustafaraj. 2009. The Battle for the 2008 US Congressional Elections on the Web. In *Proceedings of the 1st WebSci'09 Conference: Society on-line*. Athens, Greece.

[31] P Takis Metaxas and Yada Pruksachatkun. 2017. Manipulation of search engine results during the 2016 US congressional elections. In *Proceedings of the Twelfth International Conference on Internet and Web Applications and Services (ICIW)*.

[32] Thomas E. Nelson and Donald R. Kinder. 1996. Issue Frames and Group-Centrism in American Public Opinion. *The Journal of Politics* 58, 4 (1996), 1055–1078. http://www.jstor.org/stable/2960149

[33] Alexandros Ntoulas, Marc Najork, Mark Manasse, and Dennis Fetterly. 2006. Detecting spam web pages through content analysis. In *Proceedings of the 15th international conference on World Wide Web*. ACM, 83–92.

[34] Bing Pan, Helene Hembrooke, Thorsten Joachims, Lori Lorigo, Geri Gay, and Laura Granka. 2007. In Google we trust: Users' decisions on rank, position, and relevance. *Journal of computer-mediated communication* 12, 3 (2007), 801–823.

[35] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think.* Penguin.

[36] Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1650–1659.

[37] Ronald E. Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. 2018. Auditing Partisan Audience Bias Within Google Search. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 148 (Nov. 2018), 22 pages. https://doi.org/10.1145/3274417

[38] Ronald E. Robertson, Shan Jiang, David Lazer, and Christo Wilson. 2019. Auditing Autocomplete: Suggestion Networks and Recursive Algorithm Interrogation. In *Proceedings of the 10th ACM Conference on Web Science (WebSci '19)*. ACM, New York, NY, USA, 235–244. https://doi.org/10.1145/3292522.3326047

[39] Ronald E. Robertson, David Lazer, and Christo Wilson. 2018. Auditing the Personalization and Composition of Politically-Related Search Engine Results Pages. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 955–965. https://doi.org/10.1145/3178876.3186143

[40] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry* 22 (2014).

[41] Tali Sharot. 2011. The optimism bias. *Current biology* 21, 23 (2011), R941–R945.

[42] Dan Sperber and Deirdre Wilson. 2004. Relevance theory. *Handbook of Pragmatics. Oxford: Blackwell* (2004), 607–632.

[43] Kate Starbird. 2019. Disinformation's spread: bots, trolls and all of us. *Nature* 571, 7766 (2019), 449.

[44] Seth Stephens-Davidowitz. 2017. *Everybody lies: Big data, new data, and what the internet can tell us about who we really are.* HarperCollins New York.

[45] Daniel Trielli and Nicholas Diakopoulos. 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)*. ACM, New York, NY, USA, Article 453, 15 pages. https://doi.org/10.1145/3290605.3300683

[46] Francesca Tripodi. 2018. Searching for Alternative Facts: Analyzing Scriptural Inference in Conservative News Practices. *Data & Society* (2018).

[47] Nicholas A. Valentino, Vincent L. Hutchings, Antoine J. Banks, and Anne K. Davis. 2008. Is a Worried Citizen a Good Citizen? Emotions, Political Information Seeking, and Learning via the Internet. *Political Psychology* 29, 2 (2008), 247–273. https://doi.org/10.1111/j.1467-9221.2008.00625.x

[48] Liwen Vaughan and Mike Thelwall. 2004. Search engine coverage bias: evidence and possible causes. *Information Processing Management* 40, 4 (2004), 693 – 707. https://doi.org/10.1016/S0306-4573(03)00063-3

[49] Ingmar Weber, Venkata Rama Kiran Garimella, and Erik Borra. 2012. Mining web query logs to analyze political issues. In *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 330–334.