Auditing the Partisanship of Google Search Snippets

Desheng Hu
Northeastern University
hu.des@husky.neu.edu

Shan Jiang
Northeastern University
sjiang@ccs.neu.edu

Ronald E. Robertson
Northeastern University
rer@ccs.neu.edu

Christo Wilson
Northeastern University
cbw@ccs.neu.edu

ABSTRACT

The text snippets presented in web search results provide users with a slice of page content that they can quickly scan to help inform their click decisions. However, little is known about how these snippets are generated or how they relate to a user's search query. Motivated by the growing body of evidence suggesting that search engine rankings can influence undecided voters, we conducted an algorithm audit of the political partisanship of Google Search snippets relative to the webpages they are extracted from. To accomplish this, we constructed lexicon of partisan cues to measure partisanship and construct a set of left- and right-leaning search queries. Then, we collected a large dataset of Search Engine Results Pages (SERPs) by running our partisan queries and their autocomplete suggestions on Google Search. After using our lexicon to score the machine-coded partisanship of snippets and webpages, we found that Google Search's snippets generally amplify partisanship, and that this effect is robust across different types of webpages, query topics, and partisan (left- and right-leaning) queries.

CCS CONCEPTS

• **Information systems → Summarization**; • **Social and professional topics → Political speech**; • **Human-centered computing →** *User interface design.*

KEYWORDS

snippet generation; Google Search; partisan echo chamber; algorithm auditing

ACM Reference Format:
Desheng Hu, Shan Jiang, Ronald E. Robertson, and Christo Wilson. 2019. Auditing the Partisanship of Google Search Snippets. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3308558.3313654

1 INTRODUCTION

Search engines are one of, if not the most important tool used by people seeking information on the web. Surveys have repeatedly found that people reach for search engines first when they have online information needs [18] and that this happens on a daily basis [19]. These observations are especially true with respect to breaking news, where people report getting news from search engines even more frequently than social media [51, 54]. Surveys and ethnographic studies have also observed that search engines are specifically used for "fact-checking," possibly because search engines are viewed as neutral and trusted sources of information [4, 18, 19, 30, 55, 65].

When conceptualizing the role of search engines in information seeking, it is critical to recognize that they are no longer just intermediaries that transfer users from queries to websites via "10 blue links" [46]. Instead, modern search engines are rich media platforms unto themselves, often presenting links to third-party websites alongside images, videos, maps, algorithmically curated "knowledge," and social media posts [58]. This additional control over how information is presented — beyond ranking and filtering — can further increase the capacity of search engines to shape users' behavior and preferences. For example, recent studies have found that different types of queries often return different types of results, and these different result types (e.g., embedded Twitter results) can surface partisan or low-quality links in highly ranked positions [16, 57, 58].

One particularly important feature found on modern Search Engine Result Pages (SERPs) are *search snippets*. Snippets appear below links to webpages and provide a brief summary of the content of the given page. These summaries provide users with additional context about a webpage, beyond what can be gleaned from their title. Originally, snippets were drawn from HTML `meta description` tags embedded in webpages. Over time, however, search engines have incorporated *document summarization algorithms* [64] to dynamically produce snippets for pages that lack meta-data, as well as combat intentionally misleading meta-data [24, 28, 63].

Snippets are a critical — and, we argue, under-scrutinized — facet of modern search engines. *First*, snippets have a direct influence on the links that users choose to visit. For example, an inaccurate or misleading snippet may cause a user to select a different link. *Second*, even if a user does not click a given link (or *any* of the links on a SERP), they may still read and be influenced by the snippets. In much the same way people scan the titles and opening paragraphs of newspaper articles [25, 27], users may scan the titles and snippets of webpages in a SERP.

In this study, we present the first analysis of *partisan cues* in snippets on Google Search. We define partisan cues as words or phrases that are disproportionately used by a specific political group to signal group membership and frame specific issues. For example, we observe that the phrases "gun rights" and "gun violence" are used by right- and left-leaning US politicians, respectively, to

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313654

engage with their fellow partisans and frame the gun control debate. We focus on partisan cues in SERPs returned for political queries because (1) news-seeking and fact-checking are common and socially-important use cases for Google Search, and (2) behavioral experiments have shown that the political valence of search rankings can have dramatic real-world impacts on voting behavior [22, 23].

To implement our study, we trained a lexicon of partisan unigrams and bigrams drawn from speeches given by US politicians using Natural Language Processing (NLP) techniques [26]. This lexicon serves two purposes. *First*, we manually curated a subset of the bigrams to use as queries on Google Search. In total, we chose 1,050 right- and left-leaning terms, as well as 3,520 names of US politicians, to use as queries. We queried Google Search for these terms and all of their *autocomplete* suggestions between October 13–30, 2018, ultimately producing a dataset of 88,745 SERPs and 541,437 unique webpages that were linked from the SERPs. *Second*, we used the lexicon to score the partisanship of text in our dataset from left- (-1) to right-leaning (1). We then compared the partisanship score of each snippet to the score of the webpage it was drawn from to understand whether Google Search is amplifying, dampening, or even inverting the incidence of partisan cues in snippet text relative to the original document.

Our study makes the following contributions and observations:

- We present the first large-scale analysis of machine-coded partisanship in Google Search snippets, covering 4,570 political queries and their autocomplete suggestions.
- We audit the behavior of Google Search's document summarization algorithm, and find that snippets tend to be drawn from text that is near the beginning of webpages. We further observe that the algorithm leverages visible text and textual meta-data (such as alt-text on images) from webpages.
- Overall, we find that 54–58% of snippets amplify partisanship, depending on the fraction of our lexicon that is used for scoring, i.e., the snippets contain stronger partisan cues on average than the corresponding webpage they were synthesized from. This finding remains consistent across SERPs from left- and right-leaning queries and pages with and without structured meta-data that may influence Google Search's document summarization algorithm [28, 29].
- Surprisingly, we find that 19–24% of snippets have inverse partisanship than the corresponding webpage.
- We identify 31 websites where Google Search consistently produces snippets that differ from the underlying webpages in terms of the machine-coded partisanship, with high statistical significance. These websites include prominent news and social media services.

We believe that it is highly unlikely that Google has intentionally engineered their document summarization algorithm to amplify partisan cues. Instead, a more likely explanation for our findings is that journalistic practice encourages the use of partisan terms and quotes from partisan politicians in the introduction (and meta-data) of articles, which are also the types of text favored by the summarization algorithm. We hope this study will foster additional research and public debate about the role of search snippets in web search, and foster the development of document summarization algorithms that are intentionally designed to align with societal expectations and democratic norms.

Outline. Our study is organized as follows: in § 2 we motivate our study and survey related work. In § 3 we introduce our datasets and partisan scoring metrics. We analyze our dataset in § 4 and conclude with limitations and discussion in § 5.

2 BACKGROUND AND RELATED WORK

Snippets are a ubiquitous and important feature on modern search engines like Google Search. The goal of snippets is to help users assess the relevance of linked documents without having to click through each one [64, 67]. Snippets appear below links in SERPs and typically show one to three lines of text summarizing the linked document. The sentences or fragments that compose a given snippet may be extracted from different parts of source document, including visible text and invisible meta-data such as the HTML `meta description` tag, alt-text on HTML `img` tags, and various Microformat and Microdata structured meta-data languages [28, 29]. Search engines often generate snippets dynamically at query-time to select document text that is relevant to the current user's query [3, 64].

Document Summarization. Automated snippet generation is a variant of extractive summarization (in contrast to abstractive summarization [10]) that has been actively researched for decades. The earliest text summarization algorithms relied on word frequency distributions [45], sentence location [21], and per-sentence scoring techniques [52] to select text for the summary. In 1998, methods that tailored summaries to specific queries were invented [64], which continue to influence the design of modern search engines [5, 68]. More recent research has used machine learning [36, 69] and neural network-based methods [11, 50] to achieve better summarization performance.

Studying Web Search Engines. Prior work has examined various qualitative aspects of snippets. There are studies on whether the length of snippets impact user experience and the informativeness of search results [13, 38], as well as how snippet length should be customized for mobile web search [37]. Collins-Thompson et al. studied the relationship between user experience and the reading levels, and found that dwell-time improved when the reading level of the snippet matched the underlying document [12]. In our context, this suggests that user satisfaction may be maximized when the partisan cues in a snippet accurately reflect the partisanship of the corresponding document.

To our knowledge, no prior studies have examined the partisanship of snippets. Although there are several studies that have *audited* [15, 59] Google Search, the focus of these studies has been examining how Google Search's personalization algorithm impacts the links that appear in SERPs [32, 39, 58], or the partisanship of links in SERPs [16, 57].

The Impact of Snippets. Our study of partisan cues in snippets is motivated by work that examined how people read and interpret the news. *First*, studies have shown that people often do not read news articles in full; instead, they scan the headlines [25, 27]. Eye-tracking studies suggest that the same behavior is true with respect to snippets, since they attract a major portion of users' attention when they browse SERPs [13, 31], but ultimately users may only click a single link (if any) to view in full. *Second*, news headlines have the power to shape users' perceptions of stories, even if they read the full articles [20, 25]. We hypothesize that the same effect may be true of snippets: the partisan cues highlighted in a snippet may influence users' perceptions of the linked document even if it has a different partisan slant. Furthermore, in cases where a user does not click any links on a SERP, partisan cues in the snippets may still influence the user's perception of the query term itself.

Scoring Partisanship. Identifying ideological bias in text is difficult and many methods have been proposed to address it. Studies on political communication have revealed that subtle differences in word choice can signal an ideological leaning [9, 17, 35]. For example, US Republicans say "death tax" while Democrats say "estate tax," and there are no ideologically neutral alternative phrases [41].

Early approaches to political text analysis typically used human raters to manually code political statements [7, 8, 40] or relied on manually constructed dictionaries of terms to automatically label statements [14, 33, 34, 43, 53, 61, 62]. In this study, we leverage NLP techniques that automatically generate lexicons of partisan terms using politician's public speeches as a convenient, voluminous, and well-labeled (i.e., because the partisanship of the speaker can be determined based on their party affiliation and voting record) training dataset [26, 60]. These techniques obviate the need for tedious human labeling. Using the lexicons, the *machine-coded* partisanship of a document can be quantified by looking at the number and strength of polarized terms in the document [42, 44].

Unlike prior work on partisanship of SERPs, our goal is to examine the partisanship of text rather than of web domains [16, 57]. Thus, existing datasets that score the partisanship of web domains are not pertinent to our study [1, 2, 6, 49, 56].

3 DATASET AND METHODOLOGY

In this section, we present the datasets and methods that we use in our study. *First*, we introduce our lexicon of partisan unigrams and bigrams, including the raw data and methods we used to construct it. We use this lexicon to inform our selection of queries on Google Search, as well as score the partisanship of text. *Second*, we explain the process we used to select queries for Google Search, crawl SERPs and webpages, and extract data from these web documents. *Third*, we present our metrics for computing partisanship and comparing the partisanship between snippets and webpages.

3.1 Lexicon Construction

The basis for our study is a lexicon of unigrams and bigrams[1] that represent *partisan cues*, i.e., each phrase is primarily used by partisan people with a specific ideological stance. We use this lexicon for two purposes: to select ideological queries for Google Search (see § 3.2) and to score the partisanship of text (see § 3.3).

We use a method developed by Gentzkow and Shapiro to construct our lexicon [26]. In brief, this method counts the frequency of n-grams in documents that have reliable partisan labels (e.g., right and left), and then scores each n-gram based on the relative, normalized frequency of its use by individuals on each side of the partisan divide.

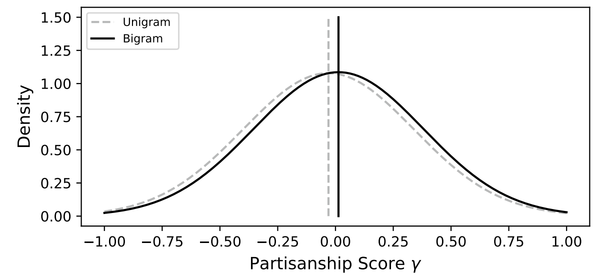


Figure 1: Distribution of partisanship scores γ of unigrams and bigrams in our Vote Smart lexicon.

For our study, we crawled a corpus of documents from votesmart.org, which is a website that catalogues information about politicians, on August 11, 2018. We chose Vote Smart because prior work has shown that its data is useful for understanding politicized language [66]. Specifically, we gathered transcripts of speeches and letters from 20 Democrats and 19 Republicans, including Donald Trump, Hillary Clinton, all congressional leaders from the US House and Senate, and several other prominent US congressional representatives. We selected this set of politicians because they generate a large volume of documents due to their public presence, and as leaders they tend to set the agendas for their respective parties.

Before constructing our lexicon, we balanced and cleaned the corpus. *First*, we chose to use documents dated between January 2008 and August 2018, as this covered the current and former US presidential administrations. We found that using older documents polluted the final lexicon with phrases and issues that were no longer salient in 2018, meaning they did not return relevant results when queried on Google Search. *Second*, we randomly downsampled the documents each year to ensure that each side had an equal number of total documents. *Third*, we cleaned (e.g., case-folding, lemmatization, etc.) and tokenized the corpus using NLTK.[2]

Next, we extracted all unigrams and bigrams from the cleaned corpus and scored them. We computed the probability of each unigram and bigram g appearing in Republican- and Democrat-authored text as $Pr_R(g)$ and $\Pr_D(g)$, respectively. We then computed the partisan bias γ of each g as

$$\gamma(g) = \frac{Pr_R(g) - Pr_D(g)}{Pr_R(g) + Pr_D(g)}. \quad (1)$$

γ scores range from $[-1, 1]$, with 1 (-1) indicating that a phrase is used exclusively by Republicans (Democrats).

Finally, we obtained our lexicon by filtering out unigrams and bigrams that appear <50 times, as these are unlikely to be recognized by people as partisan. After filtering, our lexicon contains 69,901 total terms, with 13,464 unigrams and 56,437 bigrams. Example terms and their scores include: "LGBT" (-0.92), "climate crisis" (-0.91), "equal pay" (-0.90), "gun safety" (-0.85), "burdensome regulation"

[1]We do not consider n-grams with $n > 2$ in this study because we believe that the majority of commonly used partisan cues can be expressed in one or two words.

[2]NLTK: www.nltk.org/

Table 1: Example root queries organized by topic. Queries are color coded as right- and left-leaning by partisanship score.

Topic	Root Queries
Gun	gun control, conceal weapon, gun rights, traffic firearm, gun trace, illegal gun, crime gun, gun traffic, gun dealer, gun lobby, gun safety, gun violence
Election	military voter, secret ballot, voter assistance, ballot access, voter suppression, election act, election infrastructure, election security, eligible voter, secure election, interference 2016 election
President	president health, executive overreach, pipeline president, mandate president, assure president, president unconstitutional, truth president, border president, president hillary, initiative president, president emergency, president assad
Pregnancy	elective abortion, largest abortion, abortion provider, pay abortion, perform abortion, abortion coverage, abortion clinic, fund abortion, unintended pregnancy, emergency contraception, access contraception
Tax	territorial tax system, flat tax, first 24000 tax free, tax mandate, repeal death tax, burdensome taxes, tax scam, tax giveaway, tuition tax, manufacture tax, tax wealthiest, offshore tax haven
American	burden american, regulation american, american conservative, american strength, american liberty, american patriot, american hostage, american spirit, eligible american, american caucus, wealthiest american, richest american, muslim american, american democracy, reignite american, american privacy
Statutes	scrub act, retirement act, reins act, act unilaterally, raise the wage act, honest ads act, undermine affordable care act, equality act, real id act, refinance act, second chance act, pay act
Medicare	raid medicare, medicare trustee, care entitlement, higher copays, medicare advantage, medicaid patient, medicare negotiate, affordable child, cost prescription, cut medicaid, improve affordable, benefit affordable
LGBT	lgbt rights, lgbt american, gay bisexual, lesbian gay, lgbt people, lgbt individual, lgbt community, discrimination lgbt
Obamacare	obamacare mandate, obamacare premium, obamacare plan, problem obamacare, failure obamacare, obamacare employer, obamacare break, obamacare disaster

(0.81), "illegal immigrant" (0.86), "bureaucrat" (0.89), "obamacare" (0.91), and "death tax" (0.98).

Figure 1 shows the score distribution for terms in the lexicon. As expected, the scores are normally distributed, with most terms having low partisanship. The mean score for unigrams and bigrams are -0.03 and 0.01, respectively, demonstrating that the final lexicon is well-balanced.

3.2 Data Collection and Extraction

The next step in our methodology is constructing a dataset of snippets from Google Search SERPs along with the corresponding webpages. Constructing this dataset involved choosing the queries for Google Search, crawling, and finally extracting data from the resulting webpages.

Query Selection. Our ultimate goal in this study is to compare the prevalence of partisan cues in Google Search snippets to their corresponding webpages. To achieve this, we need a large sample of politically-relevant SERPs from Google Search, which necessitates having a set of politically-relevant queries.

We built our set of *root queries* from two sources: (1) the names of US politicians and (2) partisan bigrams from our lexicon. For the former, we gathered the names of 3,520 US politicians, including prominent members of the current and former presidential administrations, members of the US House and Senate, and state governors. For the latter, we had two independent human labelers[3] examine all bigrams in the lexicon with scores < -0.5 or > 0.5 and manually curate terms that were specific enough to be valid queries on Google Search. For example, the bigrams "spend percentage" and "defend proud" both have a score of 1 in the lexicon (i.e., exclusive use by Republicans), but the labelers felt that these bigrams (and others like them) were not useful as queries, since they were unlikely to return results that linked to news and political webpages. Additionally, the labelers expanded bigrams in cases where additional words were necessary to transform them into valid queries. For example, the labelers expanded the bigrams "sanctity life" and "operation choke" (both with a score of 1) to "sanctity of life" and "operation choke point". Because this labeling task is highly subjective, we accepted all of the labelers' suggestions and then randomly downsampled to produce an ideologically balanced set of 1,050 partisan queries.

Clustering. To facilitate analysis of how partisan cues vary by topic, we clustered the 1,050 partisan root queries using a word-embedding technique [48]. Specifically, we used the pre-trained word vector *wiki-news-300d-1M.vec* [47] to map each root query into a vector space. Then we averaged the vectors corresponding to the words in each root query and clustered the mean vectors using k-means. The authors manually examined all 41 resulting clusters and selected 24 that were well-formed, i.e., all of the included root queries were related to a single, well-defined topic. Table 1 shows a subset of these topical clusters and the queries they contained.

Crawling and Query Expansion. We queried Google Search for all 4,570 root queries between October 13–30, 2018. We also queried Google Search for all of the *autocomplete suggestions* for each root query. This is a useful technique for expanding our root queries because the suggestions are drawn from real user queries, i.e., the suggestions provide a window into an more ecologically-valid sample of queries [57, 58]. Although some of the autocomplete suggestions may not be politically-relevant, the corresponding SERPs have minimal impact our analysis because they typically do not contain terms from our lexicon. In total, we ran 88,745 queries and collected an equal number of SERPs from Google Search.

Data Extraction. Since our focus is on the relationship between search snippets and webpage content, we parsed all of the SERPs in our dataset to retrieve snippets and their associated webpage links. We filtered out links in the SERPs that did not have snippets, linked to back to *.google.com, or were part of a non-standard *component* (e.g., maps, videos, images, etc.) [57, 58]. Next, we crawled all the links. Finally, for each (snippet, webpage) tuple, we filtered out cases were the webpage contained <300 characters, which removed 114,020 tuples from our dataset. Ultimately, our dataset contained 820,795 (snippet, webpage) tuples.

In most of our analysis in § 4 we focus on the *page text* in webpages. We define page text as the string content that exists between HTML tags in the body of the HTML document. Page

[3]Both labelers were US citizens by birth and were familiar with US politics.

Table 2: Lexicon size and number of (snippet, webpage) pairs that are scorable under different thresholds.

Threshold	Unigrams	Bigrams	Scorable Pairs Count	Scorable Pairs %
5%	638	2692	43311	5.28
10%	1276	5378	137067	16.70
15%	1914	8076	272745	33.23

text excludes everything in the document head, HTML tags, tag attributes, JavaScript, and Cascading Style Sheets.

3.3 Scoring Snippets and Webpages

The next step in our methodology is computing a partisan bias score for each snippet and webpage using our lexicon. Let U and B be the sets of unigrams and bigrams in our lexicon, and $\gamma(g)$ is the partisan score of n-gram g. For a given sample of text (either a snippet or webpage) produced in response to a given query, we compute its partisan bias score Γ using the following steps:

(1) Use NLTK to clean and tokenize the text and the query, producing lists of tokens T and Q, respectively.

(2) Compute $T' = \{\forall t \in T \mid t \notin Q\}$, i.e., remove all of the tokens $q \in Q$ from T. This step ensures fairness: 1,050 of our queries are explicitly partisan, and query terms are highly likely to appear in the corresponding snippets and webpages. It would be unfair to explicitly query for partisan words then use those same words to score the resulting samples.

(3) Compute $S_B = \{\forall t \in T' \mid t \in B\}$, i.e., the list of bigrams in T'. Compute $S_U = \{\forall t \in (T' - S_B) \mid t \in U\}$, i.e., list of unigrams in T' after the bigrams have been removed (to avoid double counting cases where unigrams and bigrams have overlapping words). Compute $S = S_B \uplus S_U$. This process prioritizes bigrams over unigrams since the former tend to have stronger semantic meanings [60].

(4) Compute the partisan bias score Γ of S as

$$\Gamma(S) = \frac{\sum_{s \in S} \gamma(s)}{|S|}, \quad (2)$$

i.e., the mean bias score of all n-grams found in T'.

Γ scores range from $[-1, 1]$, with -1 (1) indicating that the partisan cues in the snippet or webpage lean heavily to the left (right).

Note that we have intentionally chosen to normalize Γ scores using the total number of partisan queues in the text, rather than the total number of tokens in the text. We chose this formulation because webpage text tends to be orders of magnitude longer than snippet text. If we normalized by the number of tokens, then webpages would be guaranteed to have much lower scores than snippets, and thus it would not be reasonable to compare the scores for a snippet and the corresponding webpage.

Relative Bias Scores. In most of the analysis in § 4, our focus is on comparing the Γ scores of (snippet, webpage) pairs. To compare the magnitude and direction of the two Γ scores, we defined a *relative political bias score* Δ computed as

$$\Delta(\Gamma_s, \Gamma_w) = \begin{cases} \Gamma_s - \Gamma_w & \text{if } \Gamma_w \geq 0 \\ \Gamma_w - \Gamma_s & \text{if } \Gamma_w < 0 \end{cases} \quad (3)$$

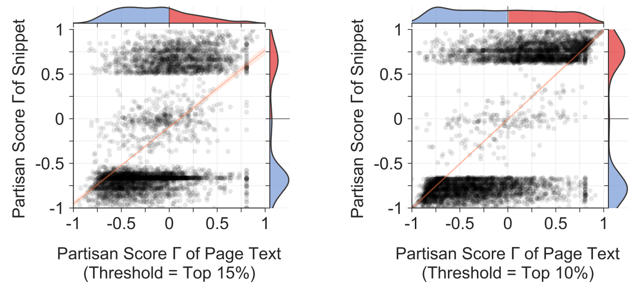


Figure 2: Partisan scores Γ of (snippet, webpage) pairs under the top 15% and top 10% lexicon thresholds.

where Γ_s and Γ_w are the Γ scores of the snippet and webpage, respectively. Δ scores exist in the range $[-2, 1]$, with three distinct subranges that have different interpretations.

- $[0, 1]$: Δ scores in this range indicate that the snippet had more partisan cues than the webpage, thus we say that the snippet *amplified* partisan bias.
- $[-2, -1]$: Δ scores in this range indicate that the snippet and webpage had opposite polarity (i.e., one was right-leaning and the other left-leaning, or vice versa), thus we say that the snippet *flipped* the partisan bias.
- $[-1, 0]$: Δ scores in this range are ambiguous. They could indicate that the snippet had fewer partisan cues than the webpage, thus the snippet *decreased* partisan bias, or they could indicate that the snippet flipped the partisan bias.

In § 4, we present both the distribution of Δ scores, as well as the percentage of cases where snippets amplified, decreased, or flipped partisan bias, to resolve ambiguity. Furthermore, to be conservative in our analysis, we also interpret Δ scores in the range $[-0.1, 0.1]$ as *unchanged*, i.e., the snippet and webpage had sufficiently similar Γ scores that a person is unlikely to notice substantive differences between their use of partisan cues.

Thresholding the Lexicon. As shown in Figure 1, the vast majority of terms in our lexicon have low γ scores, i.e., they are not strongly partisan, and are unlikely to be recognized as partisan by a human being. Since we are only interested in strongly partisan cues, we chose three thresholds $x = [5\%, 10\%, 15\%]$ to prune our lexicon. For threshold x, we select the top x left-leaning and top x right-leaning n-grams in the lexicon.

Table 2 shows the number of unigrams and bigrams that remain in the lexicon after thresholding, as well as the the number and percentage of all (snippet, webpage) pairs that are scorable at a given threshold. We say that a (snippet, webpage) pair is not scorable if Γ is undefined for either, i.e., the snippet or the webpage contains zero n-grams from the given lexicon once the words in the query itself are removed.

As expected, Table 2 shows that the percentage of our dataset that is scorable shrinks rapidly as the threshold becomes stricter. In § 4, we only analyze the datasets at the 10% and 15% thresholds because the scorable sample at 5% is too small.

4 ANALYSIS

In this section, we present our analysis comparing the partisan cues in snippets and the corresponding webpages. Throughout our analysis, we use relative political bias Δ scores to compare the partisanship of (snippet, webpage) pairs, and we present results derived from two lexicons containing the top 15% and 10% of partisan terms from our Vote Smart lexicon (see § 3.3). We begin by examining the overall distribution of Δ scores, and then examine various subsets of our full dataset to determine if Δ scores vary by search query, website, and webpage structure.

4.1 Overall Results

Figure 2 shows the Γ scores for each (snippet, webpage) pair in our dataset under the 15% and 10% lexicon thresholds. As expected, the distribution of scores for snippets is strongly bimodal because the average snippet only includes 1–2 partisan cues and they are likely to share the same polarity (e.g., both lean right or both lean left). In contrast, the distribution of scores for webpages is more evenly distributed, since the average webpage contains 30–52 partisan cues depending on the lexicon threshold. The Spearman rank correlation between the snippet's and webpage's Γ scores reveal moderate correlation with coefficients 0.50 and 0.63 at $p < 0.001$ for the top 15% and top 10% lexicon thresholds, respectively.

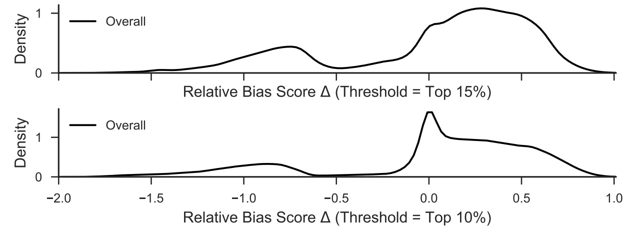
Figure 3a shows the distribution of Δ scores calculated over our entire dataset of (snippet, webpage) pairs under the top 15% and top 10% lexicons. The distributions have similar, bimodal shape regardless of lexicon: the bulk of the volume is in the $[0, 1]$ range, indicating that the snippets have stronger partisan cues than the corresponding webpages. The remaining volume tends to fall in the $[-1.5, -0.5]$, indicating snippets that either (1) have weaker partisan cues than the corresponding webpage or (2) have opposite partisan cues than the corresponding webpage (i.e., the political polarity has flipped).

The "Overall" bar in Figure 4 presents the fractions of (snippet, website) pairs where the snippet *amplified*, *decreased*, *flipped*, or had *unchanged* partisanship relative to the corresponding webpage. Recall that we conservatively consider Δ scores in the range $[-0.1, 0.1]$ to be unchanged, since the relative difference in partisanship between the snippet and webpage are unlikely to be noticeable in practice. We observe that 54–58% of snippets amplified partisan cues, depending on the lexicon threshold. Cases where polarity flipped account for 19–24% of the total, while cases where the snippet had weaker partisan cues only account for 3–4% of the total.

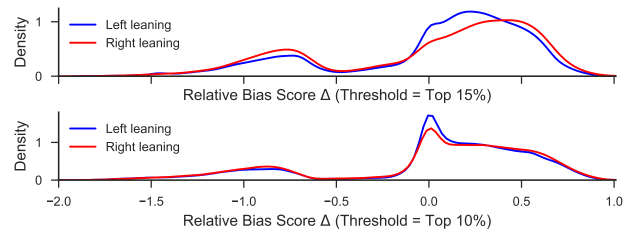
4.2 By Location

The overall results in § 4.1 surprised us: they suggest that Google Search's document summarization algorithm constructs snippets that consistently amplify partisanship. To investigate why this might be the case, we examine where Google's algorithm tends to extract snippets from in webpages, and whether this relates to the use of partisan cues.

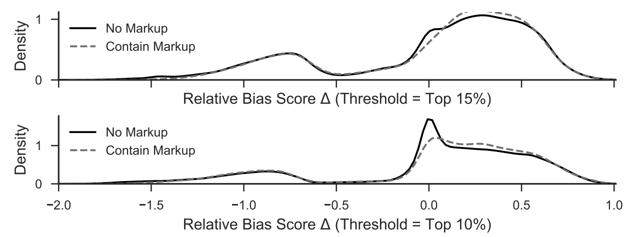
Figure 5 shows the distribution of locations within webpages that snippets were extracted from in our dataset. We split the analysis between snippets that were extracted entirely from one location, and snippets that were composed of several pieces that were extract



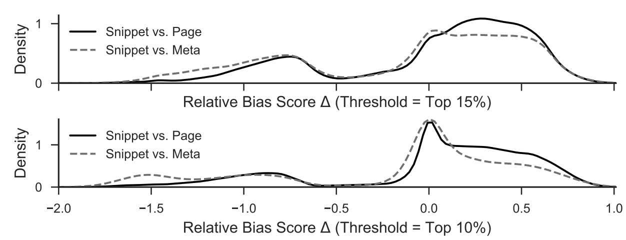
(a) All (snippet, webpage) pairs



(b) Split by query polarity



(c) Split by webpages with and without structured meta-data



(d) Snippets compared to webpage text and meta-data

Figure 3: Distribution of relative political bias scores Δ under the top 15% and top 10% lexicon thresholds. We present the overall score distribution, as well as distributions where the (snippet, webpage) pairs are split by various criteria.

from different locations in the webpage[4]. We observe that the algorithm has a strong preference for text that is earlier in documents. This may interact poorly with the "inverted pyramid" journalism

[4]In this analysis, we ignore snippets that were extracted entirely from webpage meta-data

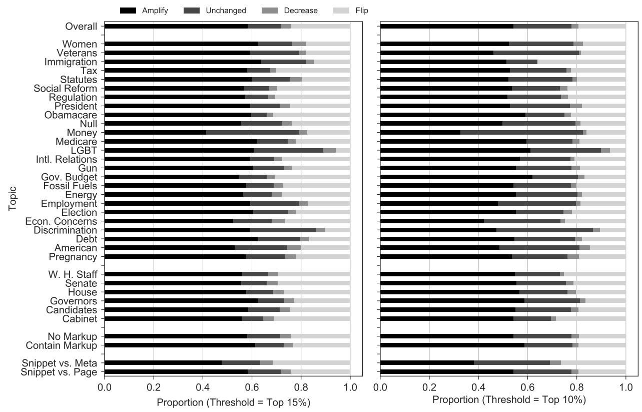


Figure 4: Fraction of cases where the snippet amplified, decreased, flipped, or had unchanged partisanship relative to the corresponding webpage. We present stacked bars for all (snippet, webpage) pairs, pairs within various query topics, pairs with webpages containing and not containing structured meta-data as well as snippets versus meta-data and versus the page text. We present results under the top 15% and top 10% lexicon thresholds.

format, which places the most specific language and jargon at the head of the story.

To further investigate the relationship between snippet location and partisanship, we calculated Δ scores between snippets and subsets of text from the corresponding webpage, where each subset contains the initial f percentage of page text. We show these results in Figure 6, varying f from 5% to 100% of the webpage text. We find that the fraction of cases where the snippet amplified the machine-coded partisanship increases as we consider larger percentages of webpage text, while the fraction of unchanged cases decreases precipitously. This suggests that either: (1) highly partisan terms at the beginning of the page, which get selected by the summarization algorithm, are eventually averaged out by less partisan terms later in the page, or (2) partisan terms of one polarity at the beginning of the page are counteracted by terms with opposite polarity later in the page. Either way, the choice to algorithmically privilege specific text snippets often leads to a mismatch between the machine-coded partisanship of the snippet versus the entire document.

4.3 By Root Query

Next, we examine how Δ scores vary by root query. To make our analysis tractable (recall that we ran 88,745 queries on Google Search), we clustered our root queries by (1) topic (see § 3.2 and Table 1) in the case of partisan phrases, and (2) job in the case of politician's names (e.g., US Senator, state governor, etc.). In total, we examine 24 topical clusters of partisan phrases and 6 clusters of politicians. Additionally, we aggregate each root query together with its associated autocomplete suggestions.

Figure 4 shows the breakdown of (snippet, webpage) relationships for different clusters of root queries. With few exceptions, the individual clusters tend to exhibit the same characteristics as the overall dataset. Notable exceptions include: the "Money" topic has

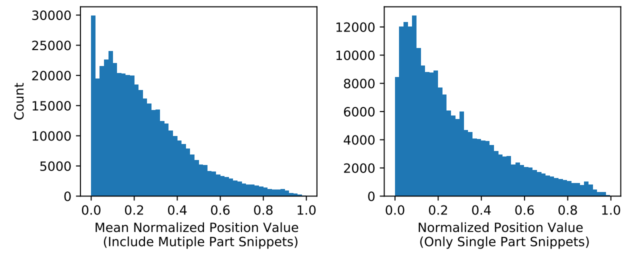


Figure 5: Distribution of the positions in webpages that snippets were extracted from, normalized over the total webpage length.

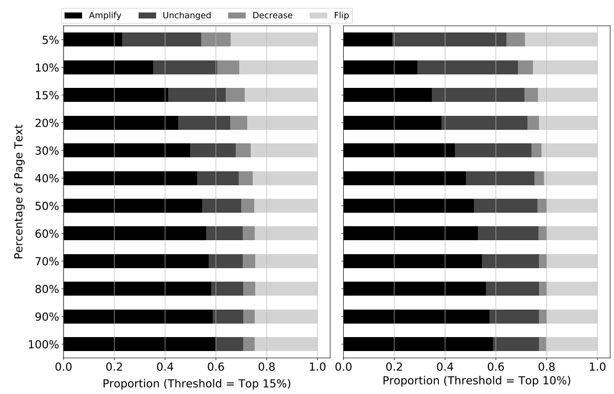


Figure 6: Fraction of cases where the snippet amplified, decreased, flipped, or had unchanged partisanship relative to subsets of text from the corresponding webpage, always starting from the top of the page. We present results under the top 15% and top 10% lexicon thresholds.

the lowest fraction of amplifying (33–41%) and highest fraction of unchanged (38–50%) snippets; under the 15% threshold, the "Immigration" and "Obamacare" topics have the largest fractions of amplifying (64%) and flipped (31%) snippets, respectively; under the 10% threshold, the "Government Budgeting" and "Immigration" topics have the largest fractions of amplifying (62%) and flipped (34%) snippets, respectively.

4.4 By Query Polarity

Next, we examine how Δ scores vary by query polarity. Specifically, we take our clusters of the root queries and divide them into left- and right-leaning queries. For partisan phrases, this is done using their γ score (see § 3.1); for politician's names, this is done using their party affiliation. Note that for this analysis, we omit six topical clusters that only contained root queries of one polarity (e.g., "LGBT", see Table 1).

Figure 3b shows the distribution of Δ scores for (snippet, webpage) pairs, split between left- and right-leaning queries. The distributions exhibit roughly the same bimodal patterns as the overall

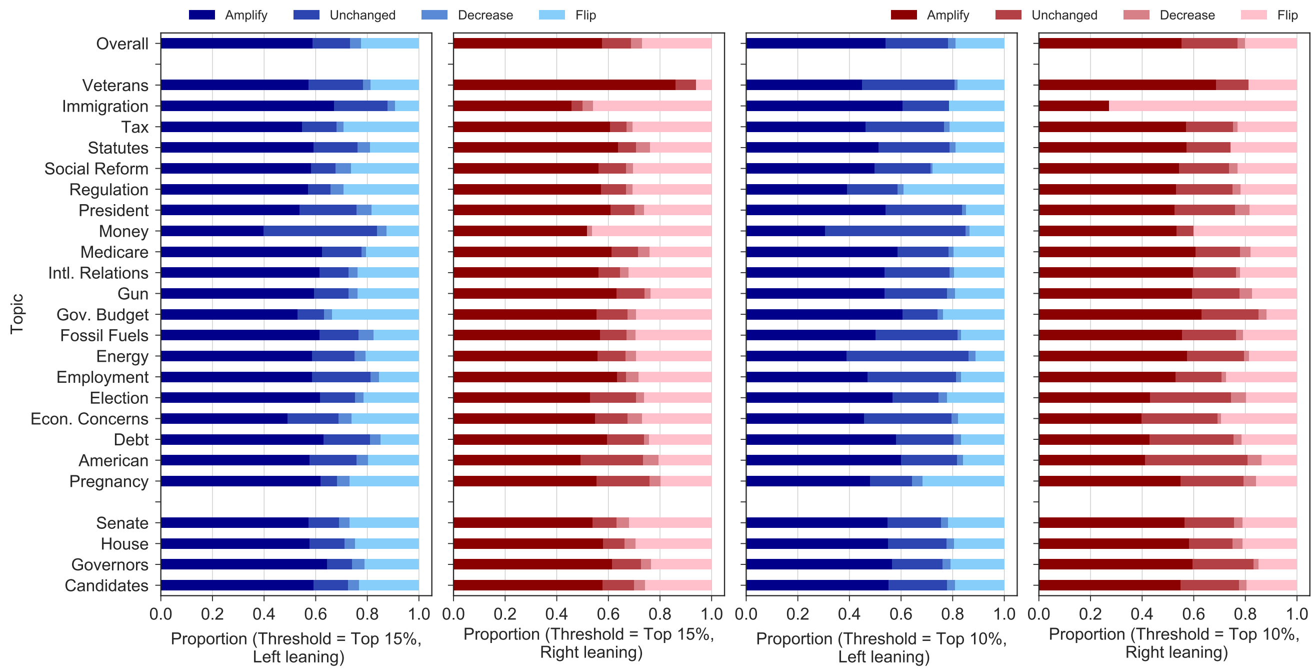


Figure 7: Fraction of cases where the snippet amplified, decreased, flipped, or had unchanged partisanship relative to the corresponding webpage, split by query polarity. We present stacked bars for all (snippet, webpage) pairs and pairs within various query topics. We present results under the top 15% and top 10% lexicon thresholds.

distributions shown in Figure 3a. With respect to each other, we used the Kolmogorov-Smirnov test to compare the left- and right-leaning distributions under each lexicon threshold, respectively, and found that they are significantly different (top 15%: $D = 0.06^{***}$; top 10%: $D = 0.03^{***}$). However, the differences between the means of the distributions are very small (top 15%: 0.02; top 10%: 0.004), meaning that the differences in the distributions have little substantive impact.

Figure 7 shows the fractions of (snippet, webpage) relationships broken down by root query and query polarity, and under different lexicon thresholds. The "Overall" bars show that left- and right-leaning queries behave similarly: 54–59% of snippets from left-leaning queries amplify partisan cues, versus 55–58% of snippets from right-leaning queries; 19–22% of snippets from left-leaning queries flip polarity, versus 20–27% of snippets from right-leaning queries. However, as we previously observed there are exceptions. For example, in the "Immigration" topic, snippets in SERPs generated by left-leaning queries amplify partisan cues 61–67% of the time, versus 27–46% of the time for right-leaning queries. Similarly, in the "Money" topic, snippets generated by left-leaning queries amplify partisan cues 30–40% of the time, versus 52–53% of the time for right-leaning queries.

4.5 By Website

Next, we analyze how Δ scores vary by website, i.e., we group the (snippet, webpage) pairs by the domain hosting the webpage. For each website that had a sample size >100, we use the Wilcoxon V signed-rank test to compare the snippets' Γ scores and webpages' Γ scores, and further applied a Boneferroni correction for multiple hypothesis testings. For the top 15% lexicon we tested 235 websites, versus 113 for the 10% threshold. Ultimately, there were 31 domains whose snippets' Γ scores were significantly different from its webpages' Γ scores under both of the lexicon thresholds. We highlight these websites in Table 3, along with the count of (snippet, webpage) pairs for each websites. We organize the websites by type, including twelve news sites, seven political websites, three social media websites, and three law-related websites. We could not categorize the last six websites.

The results in Table 3 demonstrate that there are particular websites where Google Search's summarization algorithm consistently derives snippets with different machine-coded partisanship than the webpage text. This is unsurprising in the case of social media websites, since each page may contain political content from different people, and thus a single snippet may not capture all of their views. The results are more surprising in the case of news and political websites. It is difficult to say whether these findings are caused by (1) some unanticipated, but consistent, interaction between the content or formatting on these websites and Google Search's summarization algorithm, or (2) whether this reflects an intentional effort by the authors of these websites to "optimize" their appearance on Google Search by highlighting partisan cues.

4.6 By Structured Meta-data

According to Google's documentation, its snippet generation algorithm may leverage structured meta-data from within a webpage when summarizing it [28, 29]. This meta-data can be as simple as

Table 3: Websites that have significant difference between text scores and snippet scores ***Note:*** $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Website	Wilcoxon V	Count
247sports.com	$7.78 \cdot 10^{3***}$	255
bloomberg.com	$6.56 \cdot 10^{4***}$	749
businessinsider.com	$1.59 \cdot 10^{5***}$	995
cnn.com	$3.72 \cdot 10^{5***}$	1503
dallasnews.com	$3.09 \cdot 10^{3***}$	198
espn.com	$3.50 \cdot 10^{4**}$	429
forbes.com	$3.05 \cdot 10^{5***}$	1356
fortune.com	$1.14 \cdot 10^{4**}$	259
projects.propublica.org	$2.41 \cdot 10^{3***}$	170
realclearpolitics.com	$4.26 \cdot 10^{4***}$	534
reuters.com	$5.95 \cdot 10^{4***}$	590
washingtonexaminer.com	$2.76 \cdot 10^{4***}$	397
ballotpedia.org	$2.92 \cdot 10^{6***}$	6795
congress.org	$4.28 \cdot 10^{4**}$	477
eeoc.gov	$5.37 \cdot 10^{3***}$	222
gop.gov	$4.12 \cdot 10^{3***}$	289
govtrack.us	$7.52 \cdot 10^{4***}$	626
numbersusa.com	$7.28 \cdot 10^{2***}$	141
votesmart.org	$4.54 \cdot 10^{5***}$	1901
twitter.com	$3.57 \cdot 10^{7***}$	12597
youtube.com	$2.07 \cdot 10^{5***}$	3931
yelp.com	$7.32 \cdot 10^{4***}$	679
law.cornell.edu	$3.11 \cdot 10^{3***}$	211
pview.findlaw.com	$1.71 \cdot 10^{4*}$	301
texasbar.com	$9.50 \cdot 10^{2***}$	118
beenverified.com	$3.15 \cdot 10^{4***}$	414
doctor.webmd.com	$3.33 \cdot 10^{4***}$	504
healthgrades.com	$1.22 \cdot 10^{5***}$	995
legacy.com	$1.40 \cdot 10^{6***}$	3335
ratemyprofessors.com	$7.87 \cdot 10^{3**}$	217
yellowpages.com	$2.17 \cdot 10^{4***}$	374

HTML `meta` tags, or as complex as structured meta-data markups like Microformats[5], RDFa, and Microdata. Google is intentionally ambiguous about when and whether meta-data is used to generate snippets to prevent gaming of the algorithm by Search Engine Optimization (SEO) services.

In this section, we examine whether pages that contain structured meta-data data have snippets that are more or less consistent with the webpage content. Page authors could use structured meta-data to help supply Google Search's snippet generation algorithm with information that is more representative of the webpage's content; alternatively, they could use meta-data to misrepresent the webpages content (e.g., by making it seem more partisan and inflammatory) to attract visitors.

To implement this comparison, we split our dataset into two sets of (snippet, webpage) pairs, based on whether the webpage contained structured meta-data. We specifically looked for ten kinds of Microformat meta-data, including specific attributes for objects like *People and Organizations*, *Blog Posts*, *Locations*, and *Events*. 17% of webpages in our dataset contained Microformat meta-data; of these webpages, the top objects included *People and Organizations* (62%), *Blog Posts* (30%), and *Locations* (2%).

Figure 3c shows the distribution of Δ scores for (snippet, webpage) pairs, split between webpages that do and do not contain structured meta-data. As above, we used the Kolmogorov-Smirnov test to compare the two distributions under each lexicon threshold, respectively, and found that they are significantly different (top 15%: $D = 0.04^{***}$; top 10%: $D = 0.06^{***}$). However, the differences between the means of the distributions was very small (0.03), again suggesting that the differences in the distributions has little substantive impact.

Figure 4 shows the fractions of (snippet, webpage) relationships broken down by webpages containing and not containing structured meta-data. Webpages with markup have a slightly greater fraction of snippets that amplify bias (59–61% versus 54–58%) and slightly lower fraction of snippets that leave partisanship unchanged (12–20% versus 14–24%). Based on these results, structured meta-data does not appear to make a widespread difference on the machine-coded partisanship of Google Search snippets.

Next, we examine the relationship between snippets and meta-data from a different perspective. Thus far, all of our analysis has compared the partisan cues in snippets to the partisan cues in the text of the webpage (see § 3.2). However, in practice we observe that Google Search's snippet generation algorithm often extracts sentences and phrases from meta-data in the webpage, such as the as the HTML `meta description` tag and alt-text on HTML `img` tags. This is an important distinction, because this meta-data is typically not visible to the user. This creates situations where the snippet for a webpage may be extracted from text that reader of that page may never see.

To analyze the differences between visible page text and invisible meta-data, we compare the Γ score of each snippet to the Γ scores of the page text and meta-data of the corresponding webpage.[6] To implement this comparison, we extracted all of the meta-data from the webpages in our corpus.

Figure 3d shows the distribution of Δ scores for (snippet, webpage) pairs when we compare snippets versus meta-data and versus the page text. We ran the Kolmogorov-Smirnov test on these distributions and found them to be significantly different (top 15%: $D = 0.11^{***}$; top 10%: $D = 0.04^{***}$). However, unlike all of our previous comparisons, in this case the differences between the means of the distributions were large (top 15%: 0.13; top 10%: 0.20), indicating that there are qualitative differences between meta-data and page text with respect to the use of partisan cues.

Figure 4 explores this further by presenting the fractions of cases where snippets amplified, decreased, or flipped partisanship relative to the webpage text and meta-data. When compared to meta-data, snippets were less likely to amplify partisan cues (38–48% versus 54–58%), more likely to leave partisanship unchanged (16–31% versus 14–23%), and more likely to flip polarity (26–31% versus 19–24%). There are two potential explanations for this finding: (1) Google Search's snippet generation algorithm may prefer to use phrases extracted from meta-data, thus leading to closer alignment between the partisan cues in snippets and meta-data; or (2) webpage authors may use more partisan cues in meta-data than in the page text (the average number of partisan cues in webpage meta-data in our dataset is 22–33, depending on the fraction of our lexicon that is used for scoring).

[5] Microformats: http://microformats.org/

[6] Note that the results for "Snippet vs. Page" in Figure 3d and Figure 4 are identical to the "Overall" results; we repeat them to make visual comparison easier.

5 DISCUSSION

In this work, we present the first evaluation of partisan cues in Google Search snippets. Our work relies on a dataset of 820,795 unique (snippet, webpage) pairs that were gathered from Google Search based on 88,745 unique queries for partisan phrases, names of US politicians, and all of their autocomplete suggestions. We quantify the partisan cues in our dataset using a lexicon of highly partisan unigrams and bigrams extracted from political speeches hosted by Vote Smart.

The main takeaway of our study is that Google Search snippets tend to amplify partisan cues relative to the original webpages (see Figure 4). This may have serious implications for how people consume political information from Google Search: if a person "scans the headlines" on a SERP, they may be left with a more partisan impression than if they read the source documents [25, 27]. Furthermore, partisan cues in snippets may prime users with a particular framing that will color their perception of the webpages if they click through and read them [20, 25]. Field-studies should be conducted to determine whether these hypothesized effects actually hold true in practice.

One important, unresolved question is *why* Google Search's snippet generation algorithm tends to produce snippets that amplify partisan cues. It is extremely unlikely that this is an intentional design decision on Google's part; rather, this is likely an unintended, emergent effect. Our analysis of the behavior of Google Search's snippet generation algorithm in § 4.2 and § 4.6 suggests that the cause may be interactions between writing style and webpage structure: the snippet algorithm prefers text at the beginning of pages and within meta-data, while page authors tend to use more partisan cues in these locations. Additional algorithm auditing work should be conducted to more fully explore the implementation of Google Search's snippet generation algorithm.

A second, equally important question is: is Google's current snippet generation algorithm acceptable, or should it be altered (and if so, how)? Generating partisan snippets is not necessarily bad design, if the objective is to make the partisan polarity of webpages that appear in SERPs very clear to users before they click the links. Arguably, altering the snippet generation algorithm to eschew or replace partisan cues will lead to snippets that fail to accurately reflect the content of the underlying page, which may worsen the user experience if it leads to users being unsatisfied with pages they visit. That said, it is unclear whether the snippet is the best place to implicitly convey the partisanship of webpages, since the domain of the website may already be sufficient (e.g., US users recognize the differences between Fox News and MSNBC, irrespective of any particular article).

Conversely, it can be argued that reducing or eliminating partisan cues in snippets may have overall benefits for democratic society, i.e., by reframing political hot-topics using non-partisan language that does immediately trigger framing effects and confirmation bias in readers. This could be accomplished by shifting from extractive to abstractive summarization. This design change may be especially important with respect to Google Search, since it is widely believed to be neutral, and it is the first place most web users go to investigate news and politics [4, 18, 19, 30, 55, 65]. We argue that Google, and other search engines, should engage in a transparent and deliberate effort to design how their summarization algorithms interact with news and political content, rather than allowing the algorithm to blindly function as-is, without reflection or critique.

Applications. Our methods for scoring partisan cues in text have applications beyond measurement and auditing. *First*, our methods could be used to help improve the quality of document summarization algorithms, e.g., by identifying snippets that are more closely aligned with the overall political valence of a document. *Second*, our methods could be used to inform the design of document summarization algorithms that are specifically tuned to increase diversity, e.g., by finding document with a particular political valence, or by intentionally extracting snippets from documents that have reduced or flipped polarity.

Limitations. Our study has several limitations. *First*, the Γ scores that we calculate for individual webpages are not strongly correlated with the overall partisanship of the associated websites, as measured using existing scales like AllSides and others [1, 2, 6, 57]. This is expected: the webpages we sample from any given website are not necessarily representative of the website as a whole.

Second, our approach for parsing text from webpages is overly permissive: text that is not relevant for our analysis (e.g., navigation bars, page footers, etc.) will also be extracted and included in our analysis. In practice, there is no perfect tool for extracting "content" from webpages. We do not expect this additional text to have a substantive impact on our analysis.

Third, our method for scoring partisan cues does not consider contextual semantics. For example, we do not consider negation or sarcasm. Unfortunately, this is a fundamental limitation of current NLP techniques that impacts many contemporaneous studies [57].

Fourth, our lexicon may be incomplete, e.g., there may be partisan cues that are used in the media but were not repeated sufficiently by politicians for them to emerge in our lexicon. The potential for false negatives informed our decision not to analyze (snippet, webpage) pairs containing zero partisan cues from our lexicon, since it would have been presumptuous to conclude that these samples contained "zero partisanship."

Fifth, our study is not longitudinal, and it is unclear whether our results generalize over time.

Finally, people may not recognize all of the *n*-grams in our lexicon as partisan, or they may not react to them in proportion to their γ score (i.e., people may judge them to be more partisan or less partisan). Extrapolating further, this means that people may not make the same judgments about partisan cues in snippets and webpages that we do in our analysis. Additional fieldwork is necessary to validate the *n*-grams in our lexicon and their scores, as well as evaluate people's perceptions of partisan cues in situ on SERPs.

ACKNOWLEDGMENTS

This research was supported in part by NSF grant IIS-1553088. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

[1] AllSides. 2018. Media Bias Ratings. AllSides. https://www.allsides.com/media-bias/media-bias-ratings

[2] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. Science 348, 6239 (2015), 1130–1132.

[3] Lorena Leal Bando, Falk Scholer, and Andrew Turpin. 2010. Constructing Query-biased Summaries: A Comparison of Human and System Generated Snippets. In Proceedings of the Third Symposium on Information Interaction in Context (IIiX '10). ACM, New York, NY, USA, 195–204. https://doi.org/10.1145/1840784.1840813

[4] Edelman Berland. 2017. 2017 Edelman Trust Barometer. http://www.edelman.com/trust2017/ Accessed: 2017-03-07.

[5] Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, and Jeffrey C Reynar. 2011. Phrase based snippet generation. US Patent 8,010,539.

[6] Ceren Budak, Sharad Goel, and Justin M Rao. 2016. Fair and balanced? Quantifying media bias through crowdsourced content analysis. Public Opinion Quarterly 80, S1 (2016), 250–271.

[7] Ian Budge, Hans-Dieter Klingemann, et al. 2001. Mapping policy preferences: estimates for parties, electors, and governments, 1945-1998. Vol. 1. Oxford University Press on Demand.

[8] Ian Budge, David Robertson, Derek Hearl, et al. 1987. Ideology, strategy and party change: spatial analyses of post-war election programmes in 19 democracies. Cambridge University Press.

[9] Jonathan Charteris-Black. 2006. Politicians and Rhetoric: The Persuasive Power of Metaphor.

[10] Wei-Fan Chen, Matthias Hagen, Benno Stein, and Martin Potthast. 2018. A User Study on Snippet Generation: Text Reuse vs. Paraphrases. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. ACM, 1033–1036.

[11] Jianpeng Cheng and Mirella Lapata. 2016. Neural Summarization by Extracting Sentences and Words. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Berlin, Germany, 484–494.

[12] Kevyn Collins-Thompson, Paul N Bennett, Ryen W White, Sebastian De La Chica, and David Sontag. 2011. Personalizing web search results by reading level. In Proceedings of the 20th ACM international conference on Information and knowledge management. ACM, 403–412.

[13] Edward Cutrell, Zhiwei Guan, and Ed Cutrell. 2007. What are you looking for? An eye-tracking study of information usage in web search. , 407–416 pages. https://www.microsoft.com/en-us/research/publication/what-are-you-looking-for-an-eye-tracking-study-of-information-usage-in-web-search/

[14] Miranda De Vries, Daniela Giannetti, and Lucy Mansergh. 2001. Estimating policy position from the computer coding of political texts. Estimating the policy positions of political actors (2001), 193–216.

[15] Nicholas Diakopoulos. 2014. Algorithmic accountability reporting: On the investigation of black boxes. Tow Center for Digital Journalism, Columbia University (2014).

[16] Nicholas Diakopoulos, Daniel Trielli, Jennifer Stark, and Sean Mussenden. 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.). 22.

[17] Daniel Diermeier, Jean-François Godbout, Bei Yu, and Stefan Kaufmann. 2012. Language and ideology in Congress. British Journal of Political Science 42, 1 (2012), 31–55.

[18] William H. Dutton and Grant Blank. 2013. Cultures of the Internet: The Internet in Britain: Oxford Internet Survey 2013 Report. Technical Report. Oxford Internet Institute, University of Oxford. http://oxis.oii.ox.ac.uk/wp-content/uploads/2014/11/OxIS-2013.pdf

[19] William H. Dutton, Bianca Christin Reisdorf, Elizabeth Dubois, and Grant Blank. 2017. Search and Politics: A Cross-National Survey. (2017).

[20] Ullrich K. H. Ecker, Stephan Lewandowsky, Ee Pin Chang, and Rekha Pillai. 2014. The effects of subtle misinformation in news headlines. Journal of Experimental Psychology: Applied 20, 4 (2014), 323–335.

[21] H. P. Edmundson. 1969. New Methods in Automatic Extracting. J. ACM 16, 2 (April 1969), 264–285. https://doi.org/10.1145/321510.321519

[22] 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), E4512–E4521.

[23] Robert Epstein, Ronald E. Robertson, David Lazer, and Christo Wilson. 2017. Suppressing the Search Engine Manipulation Effect (SEME). Proceedings of the ACM: Human-Computer Interaction 1, 2 (November 2017).

[24] Vanessa Fox. 2009. Google Search Now Supports Microformats and Adds "Rich Snippets" to Search Results. Search Engine Land. https://searchengineland.com/google-search-now-supports-microformats-and-adds-rich-snippets-to-search-results-19055.

[25] Renae Franiuk, Jennifer L. Seefelt, and Joseph A. Vandello. 2008. Prevalence of Rape Myths in Headlines and Their Effects on Attitudes Toward Rape. Sex Roles 58, 11–12 (June 2008), 790–801.

[26] Matthew Gentzkow and Jesse M Shapiro. 2010. What drives media slant? Evidence from US daily newspapers. Econometrica 78, 1 (2010), 35–71.

[27] M. Glenski, C. Pennycuff, and T. Weninger. 2017. Consumers and Curators: Browsing and Voting Patterns on Reddit. IEEE Transactions on Computational Social Systems 4, 4 (Dec 2017), 196–206.

[28] Google. 2018. Customizing Results Snippets. Google Custom Search Developer Documentation. https://developers.google.com/custom-search/docs/snippets.

[29] Google. 2018. Providing Structured Data. Google Custom Search Developer Documentation. https://developers.google.com/custom-search/docs/structured_data.

[30] Jeffrey Gottfried and Elisa Shearer. 2016. News use across social media platforms. Pew Research Center. http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

[31] Laura A. Granka, Thorsten Joachims, and Geri Gay. 2004. Eye-tracking Analysis of User Behavior in WWW Search. In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '04). ACM, New York, NY, USA, 478–479. https://doi.org/10.1145/1008992.1009079

[32] Aniko Hannak, Piotr Sapieżyński, Arash Molavi Kakhki, Balachander Krishnamurthy, David Lazer, Alan Mislove, and Christo Wilson. 2013. Measuring Personalization of Web Search. In WWW.

[33] Roderick P Hart. 1984. Verbal style and the presidency: A computer-based analysis. Academic Pr.

[34] Roderick P Hart. 2001. Redeveloping DICTION: theoretical considerations. Progress in communication sciences (2001), 43–60.

[35] Roderick P Hart, Jay P Childers, and Colene J Lind. 2013. Political tone: How leaders talk and why. University of Chicago Press.

[36] Mikael Kågebäck, Olof Mogren, Nina Tahmasebi, and Devdatt Dubhashi. 2014. Extractive Summarization using Continuous Vector Space Models. In Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC). Association for Computational Linguistics, 31–39. https://doi.org/10.3115/v1/W14-1504

[37] Michael Kaisser, Marti A Hearst, and John B Lowe. 2008. Improving search results quality by customizing summary lengths. Proceedings of ACL-08: HLT (2008), 701–709.

[38] Jaewon Kim, Paul Thomas, Ramesh Sankaranarayana, Tom Gedeon, and Hwan-Jin Yoon. 2017. What snippet size is needed in mobile web search?. In Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval. ACM, 97–106.

[39] 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 Internet Measurement Conference (IMC 2015). Tokyo, Japan.

[40] Hans-Dieter Klingemann, Richard I Hofferbert, and Ian Budge. 1994. Parties, policies, and democracy. Westview Press.

[41] George Lakoff. 2010. Moral politics: How liberals and conservatives think. University of Chicago Press.

[42] Michael Laver, Kenneth Benoit, and John Garry. 2003. Extracting policy positions from political texts using words as data. American Political Science Review 97, 2 (2003), 311–331.

[43] Michael Laver and John Garry. 2000. Estimating policy positions from political texts. American Journal of Political Science (2000), 619–634.

[44] Wei-Hao Lin, Eric Xing, and Alexander Hauptmann. 2008. A joint topic and perspective model for ideological discourse. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, 17–32.

[45] H. P. Luhn. 1958. The Automatic Creation of Literature Abstracts. IBM J. Res. Dev. 2, 2 (April 1958), 159–165. https://doi.org/10.1147/rd.22.0159

[46] Harry McCracken. 2016. How Google Search Got Past "10 Blue Links". Fast Company. https://www.fastcompany.com/3060044/how-google-search-got-past-10-blue-links.

[47] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. 2018. Advances in Pre-Training Distributed Word Representations. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018).

[48] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems. 3111–3119.

[49] Amy Mitchell, Jeffrey Gottfried, Jocelyn Kiley, and Katerina Eva Matsa. 2014. Political Polarization & Media Habits. Pew Research Center's Journalism Project.

[50] Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents. In AAAI. AAAI Press, 3075–3081.

[51] Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, David A. L. Levy, and Rasmus Kleis Nielsen. 2017. Reuters Institute Digital News Report 2017. Technical Report. Reuters Institute for the Study of Journalism.

[52] C. D. Paice. 1990. Constructing Literature Abstracts by Computer: Techniques and Prospects. Inf. Process. Manage. 26, 1 (April 1990), 171–186. https://doi.org/10.1016/0306-4573(90)90014-S

[53] PAUL PENNINGS and JAN KLEINNIJENHUIS. 2003. Measurement of party positions on the basis of party programmes, media coverage and voter perceptions. In Estimating the policy position of political actors. Routledge, 182–202.

[54] Media Insight Project. 2014. The Personal News Cycle: How Americans Choose to Get Their News. Media Insight Project. https://www.americanpressinstitute.org/publications/reports/survey-research/how-americans-get-news/

[55] Kristen Purcell, Joanna Brenner, and Lee Rainie. 2012. Search engine use 2012. Technical Report. Pew Research Center's Internet and American Life Project.

[56] Filipe N. Ribeiro, Lucas Henrique, Fabricio Benevenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babei, and Krishna P. Gummadi. 2018. Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale. In Proceedings of the AAAI Intl. Conference on Web and Social Media (ICWSM). Stanford, USA.

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

[58] 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 on World Wide Web. International World Wide Web Conferences Steering Committee, 955–965.

[59] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. Auditing algorithms: Research methods for detecting discrimination on internet platforms. In Proceedings of "Data and Discrimination: Converting Critical Concerns into Productive Inquiry", a Productivereconference at the 64th Annual Meeting of the International Communication Association.

[60] Yanchuan Sim, Brice DL Acree, Justin H Gross, and Noah A Smith. 2013. Measuring ideological proportions in political speeches. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 91–101.

[61] Philip J Stone, Robert F Bales, J Zvi Namenwirth, and Daniel M Ogilvie. 1962. The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. Behavioral Science 7, 4 (1962), 484–498.

[62] Philip J Stone, Dexter C Dunphy, and Marshall S Smith. 1966. The general inquirer: A computer approach to content analysis. (1966).

[63] Danny Sullivan. 2008. Google Testing Enhanced Listings, "Pagelinks" & Auto-Spelling Correction. Search Engine Land. .

[64] Anastasios Tombros and Mark Sanderson. 1998. Advantages of Query Biased Summaries in Information Retrieval. In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '98). ACM, New York, NY, USA, 2–10. https://doi.org/10.1145/290941.290947

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

[66] Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Vol. 1. 1629–1638.

[67] Andrew Turpin, Yohannes Tsegay, David Hawking, and Hugh E. Williams. 2007. Fast Generation of Result Snippets in Web Search. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '07). ACM, New York, NY, USA, 127–134. https://doi.org/10.1145/1277741.1277766

[68] Alexandre A Verstak and Anurag Acharya. 2012. Generation of document snippets based on queries and search results. US Patent 8,145,617.

[69] Kam-Fai Wong, Mingli Wu, and Wenjie Li. 2008. Extractive Summarization Using Supervised and Semi-supervised Learning. In Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1 (COLING '08). Association for Computational Linguistics, Stroudsburg, PA, USA, 985–992. http://dl.acm.org/citation.cfm?id=1599081.1599205