Measuring Political Personalization of Google News Search

Huyen Le
Department of Computer Science
University of Iowa

Raven Maragh
Department of Communication Studies
Gonzaga University

Brian Ekdale
School of Journalism and Mass Communication
University of Iowa

Andrew High
Department of Communication Arts and Sciences
Pennsylvania State University

Timothy Havens
Department of Computer Science
University of Iowa

Zubair Shafiq
Department of Computer Science
University of Iowa

ABSTRACT

There is a growing concern about the extent to which algorithmic personalization limits people's exposure to diverse viewpoints, thereby creating "filter bubbles" or "echo chambers." Prior research on web search personalization has mainly reported location-based personalization of search results. In this paper, we investigate whether web search results are personalized based on a user's browsing history, which can be inferred by search engines via third-party tracking. Specifically, we develop a "sock puppet" auditing system in which a pair of fresh browser profiles, first, visits web pages that reflect divergent political discourses and, second, executes identical politically oriented Google News searches. Comparing the search results returned by Google News for distinctly trained browser profiles, we observe statistically significant personalization that tends to reinforce the presumed partisanship.

CCS CONCEPTS

• **General and reference → Measurement**; • **Information systems → Personalization**.

KEYWORDS

Measurement; Personalization; Google; News; Search

ACM Reference Format:
Huyen Le, Raven Maragh, Brian Ekdale, Andrew High, Timothy Havens, and Zubair Shafiq. 2019. Measuring Political Personalization of Google News Search. In *Proceedings of the 2019 World Wide Web Conference (WWW'19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308558.3312504

1 INTRODUCTION

When a person looks up information on a search engine, the search does not occur in a vacuum. Before any particular search is made, the person has already lived an active digital life—reading news stories, liking pages on online social media, and posting comments on blogs—activities that are tracked across the Web to train personalization algorithms. In the current divisive political climate [9], people might be more likely to access content online that conforms to their political ideologies, but little is known about whether accessing certain political content impacts personalization of other elements of a person's online life. In this paper, we ask: *If a person's digital life demonstrates a political bias, does that person receive personalized search results and do those search results reflect that person's bias?*

Researchers have long been concerned about the effect of selective exposure—seeking information that reinforces preexisting beliefs while avoiding other information—on democratic societies [38, 43]. Through selective exposure, people with biases seek out content that conforms with their preexisting beliefs. For example, selective exposure to conservative news is associated with support for strict immigration policies [44], and people who score high on a scale of modern racism are more likely to view non-traditional Internet sources as credible sources of news [26]. Researchers have recently started to question whether algorithms create distinct personalized experiences for users, leading to so-called "filter bubbles" or "echo chambers" [7, 19, 31]. Algorithms reflect the societies in which they are produced, so it is unsurprising that they are encoded with biases [13]. For example, personalization algorithms have been found to discriminate against women [11] and people of color [29].

While selective exposure requires deliberate acts of media choice, algorithmic personalization interprets past behavior as precedent for future preference. In other words, algorithmic personalization can intensify selective exposure beyond a person's choice, resulting in a vicious cycle that can contribute to an increasingly polarized society. According to Pew Research, Americans are more ideologically polarized now than at any point in the last two decades [9]. This increase in polarization coincides with the rise of search engines and social media sites as primary sources of information. Thus, it is important to study the role of personalization algorithms employed by search engines in reinforcing pre-existing biases.

Prior research has reported significant personalization in online advertising based on a user's browsing history. Personalization in online advertising [8, 11, 37, 41] is not surprising because behavioral targeting is its key design feature. To target ads, advertisers employ sophisticated online tracking techniques for profiling users' browsing habits across the Web [12]. Naturally, we would expect search engines to leverage browsing history information obtained by online tracking for search personalization as well. However, prior studies [16, 20, 42] have reported significant search personalization mainly based on location, not browsing history.

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

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3312504

In this paper, we set out to study search personalization in Google News using a "sock puppet" auditing system [36] in which we use automated programs to impersonate users with different political ideologies. While controlling for other factors, we train a pair of fresh browser profiles by visiting websites that reflect pro-immigration and anti-immigration stances. We show that Google can infer browsing histories of our trained profiles through its pervasive online tracking network and can leverage it to personalize search results based on recent changes in its privacy policy [2, 6]. We then execute search queries on Google News related to a variety of political topics. We analyze the search results to quantify the magnitude and direction of personalization. While personalization varies depending on the search term, overall we are able to conclude that profiles trained by browsing websites reflecting distinct political positions indeed receive significant personalization that tends to reinforce the presumed partisanship.

We highlight the key contributions of our work as follows.

• **Methodology:** We expand on previous work (e.g., [15, 16]) by refining the sock puppet methodology for auditing algorithmic personalization. Our sock puppet system is designed to reflect the behavior of Twitter accounts with opposing political views. Specifically, we train browser profiles by visiting URLs extracted from the timelines of these Twitter accounts. We then search Google News simultaneously from the trained browser profiles using different search terms related to popular political issues such as immigration and foreign policy.

• **Findings:** Our comparative analysis of the search results returned by Google News provides evidence of personalization based on browsing history. Specifically, while personalization varies across different search terms, we note that the search terms receiving most personalization tend to get personalized results that reinforce the presumed partisanship. We believe that our findings are different from those reported in prior work on personalization in part because of our refined sock puppet system design.

Our work presents empirical evidence of political personalization on Google News based on browsing history. The results not only set the baseline for search personalization based on political biases in browsing history, but also contribute to the broader understanding of selective exposure and algorithmic personalization.

2 SOCK PUPPET AUDITING SYSTEM

The overarching goal of our study is to examine the extent to which a user's web browsing history affects algorithmic personalization in web search. We seek to understand both whether algorithmic personalization exists and whether it reinforces or dampens the preferences displayed in web browsing history. To this end, we design and implement a sock puppet auditing system to train fresh browser profiles by visiting hyperlinks that embody distinct political discourses and then compare Google News search results for identical politically oriented search terms. Figure 1 depicts the four main components of our system that are discussed below.

2.1 Identifying Partisan Hyperlinks

We aim to evaluate Google News personalization based on browsing histories that reflect different stances on the topic of immigration. We focus on immigration because it was a key campaign issue in the recent 2016 U.S. presidential election [5, 23]. Our goal was to find a list of websites that reflect pro- and anti-immigration stances in the context of U.S. politics. To this end, we rely on two popular Twitter accounts that reflect both discourse communities. Discourse communities are groups of people that come together, whether physically or virtually, for the purpose of building community through shared goals and forms of speech [39]. For the anti-immigration stance, we use @wginfonetorg, which belongs to a popular anti-immigrant website whitegenocide.info that is dedicated to "fighting the crime of white genocide." For the pro-immigration stance, we use @DefineAmerican, which belongs to a popular pro-immigrant website defineamerican.com that is dedicated to "shift the conversation about immigrants, identity, and citizenship in a changing America." Both of these Twitter accounts are very active and fairly popular. As listed in Table 1, @wginfonetorg has posted more than 63K tweets and has 9.4K followers while @DefineAmerican has posted more than 17K tweets and has 30.3K followers. Because both accounts frequently share content from other websites that support their respective views on immigration, we collected the hyperlinks posted on the timelines of these Twitter accounts over the duration of two weeks in March 2017. We treat these two sets of hyperlinks as representations of pro- and anti-immigration discourses.

Screen Name	Number of Favorites	Number of Followers	Number of Followings	Number of Tweets	Account Creation
@DefineAmerican	10,043	30,253	1,776	17,350	May 2011
@wginfonetorg	1,253	9,424	223	63,916	April 2014

Table 1: Statistics of pro-immigration and anti-immigration Twitter accounts.

2.2 Training Browser Profiles

Using the two sets of hyperlinks collected from pro- and anti-immigration Twitter accounts, we now train fresh browser profiles. Specifically, to train the pro-immigration browser profile, we install a fresh copy of Firefox web browser and open hyperlinks from the timeline of the pro-immigration Twitter account. Similarly, to train the anti-immigration browser profile, we install a fresh copy of Firefox web browser and open hyperlinks from the timeline of the anti-immigration Twitter account. We take several steps to imitate a real user during browser profile training. First, we limit the number of hyperlinks that we use for training per day to 50. To this end, we randomly sample a subset of 50 hyperlinks from the hyperlinks posted on the timelines of both Twitter accounts. Second, we add a random delay averaged at five minutes between opening consecutive hyperlinks during training. Finally, we conduct the training between 9 am and 11 pm local time to reflect diurnal user activity.

It is noteworthy that we do not create or log-in to a Firefox account during training because it can save bookmarks, passwords, browsing history, and cookie information. We also do not create or log-in to a Google account that may be used by Google to link our account information with browsing history. Therefore, the only way for Google to know about our browsing history during training would be via third-party tracking using its own advertising/analytics network (e.g., DoubleClick and Google Analytics) via cookies or browser fingerprinting [12, 24]. As discussed in related work, prior research has reported on the extensive third-party tracking capabilities of Google.

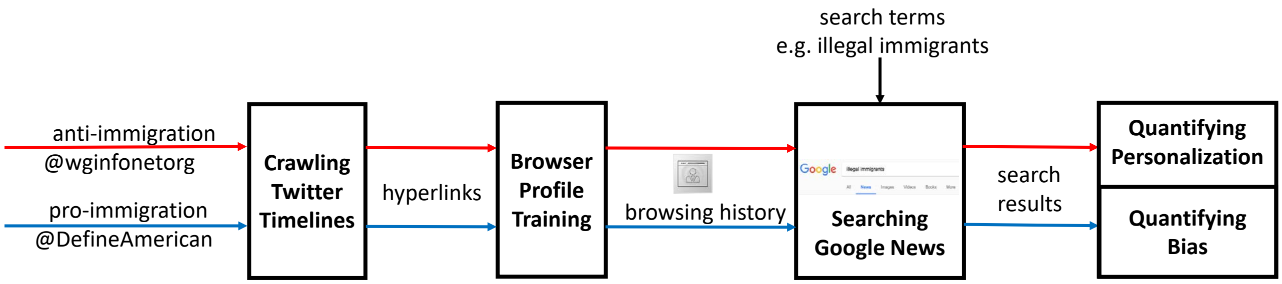

Figure 1: Our sock puppet auditing system to measure political personalization of Google News search.

Our analysis of the set of hyperlinks used for training browser profiles shows that pro-immigration training hyperlinks belong to more mainstream news outlets such as Washington Post while anti-immigration training hyperlinks belong to more alternative news sites such as Breitbart and user-generated content sites such as YouTube and WordPress. We find that several Google owned domains are top third-parties on these training hyperlinks. For example, doubleclick.net is the top third-party tracker on both anti- and pro-immigration training hyperlink sets. Thus, we conclude that Google is able to learn the browsing histories of our trained pro- and anti-immigration browser profiles and *can* later use it for personalization of search results.

2.3 Searching Google News

After training fresh browser profiles using hyperlinks crawled from pro- and anti-immigration Twitter accounts, we are set to conduct Google News searches. We are interested in testing if web browsing histories that reflect these divergent discourses would result in search personalization along politically partisan lines.[1] In other words, we want to test whether a user who consistently consumes anti-immigration content would receive more or less Republican-leaning news stories in the search results. We execute searches from three sets of browser profiles: (1) pro-immigration, (2) anti-immigration, and (3) control. The control browser profile, which is not trained (i.e., by visiting hyperlinks), provides us a “neutral” perspective to judge whether personalization occurs for pro- and/or anti-immigration profiles. We execute Google News searches using different search terms about a wide variety of political topics. We decide to search for news related to the top-10 most discussed policy issues on Twitter [23] during the 2016 U.S. presidential election campaign. We use five different search terms for each of these policy issues: Immigration, Foreign Policy, Healthcare, Economy, Abortion, Gay Rights, Gun Control, Climate Change, Education, and Veterans. For each search term, we repeat the search process every day over the duration of one week.

To ensure that the search results returned by Google are not impacted by anything other than the browsing histories, our search process is designed as follows. First, we use Selenium WebDriver [1] to automatically conduct searches using the Google News web interface instead of Google’s search API. We configure Google News to return up to 100 search results. Second, we use a wait period of 11 minutes between consecutive searches to eliminate any carry-over effect [16]. Third, we conduct searches from pro-immigration, anti-immigration, and control browser profiles in a synchronized fashion to eliminate any temporal effects. Fourth, each browser profile conducts search queries on a separate Amazon EC2 cloud instance to avoid any interdependencies between different profiles.[2] Fifth, we use a static DNS entry for Google News to ensure that our search queries are routed to the same datacenter. Finally, to eliminate potential noise in search results due to A/B testing [30], we train four separate profiles for pro-immigration, anti-immigration, and control browser profiles. In the absence of A/B testing, we would expect the same search results for these four profiles. When there is A/B testing, we can mitigate its effect by eliminating search result differences. To this end, we compute the pairwise intersection of search results among the four profiles. We then use search results from a randomly selected profile from the pair with the maximum intersection.

2.4 Quantifying Search Personalization

We use trained browser profiles for both pro- and anti-immigration stances as well as a control browser profile to execute Google News searches. We can quantify personalization by comparing the presence and ranking of search results across different browser profiles. First, we quantify personalization simply in terms of the differences in search results. Let P and A respectively denote the lists of search results for the pro- and anti-immigration browser profiles. Let C denote the list of search results for the control browser profile. We can measure the differences in search results by comparing P, A, and C in a pairwise manner. To this end, let $P - A$ represent the set of unique search results for the pro-immigration profile that do not appear for the anti-immigration profile. Let $P - C$ and $A - C$ respectively represent the set of unique search results for the pro-immigration and anti-immigration profiles with respect to the control. We use $|P - A|$, $|P - C|$, and $|A - C|$ to quantify the differences in search results in the range of [0%,100%], where 0% indicates no personalized search results and 100% indicates all 100 search results are personalized. Note that since we get 100 search results for each search term, we have $|P| = |A| = 100$. In this case,

[1] Note that the extreme anti-immigration position of those concerned about white genocide are not wholly aligned with the Republican Party’s nor are the political views of our pro-immigration profile identical to the Democratic Party platform.

[2] Note that while each Amazon cloud instance has a different IP address, they all belong to the same /24 IP prefix range geolocated in Oregon. Therefore, we do not expect any IP geolocation based search personalization as has been reported in prior work [16, 20].

the difference in search results is symmetric in terms of quantities (i.e. $|P - A| = |A - P|$) although it may be not be symmetric as the set (i.e. $P - A \neq A - P$). Second, we quantify personalization while taking into account the ranking of search results across different browser profiles. To this end, we measure edit distance among P, A, and C in a pairwise manner. Specifically, we compute the Damerau-Levenshtein distance [10] which represents the number of deletions, insertions, or substitutions needed to make a pair of lists identical. If two lists are identical, the edit distance is 0. The greater the edit distance, the more different the lists are in terms of their membership and ordering. Note that both the difference and edit distance are calculated based on search results as URLs and we do not consider title or text of these URLs.

In addition to quantifying personalization, we also analyze whether personalized search results reflect political bias. To assess political bias, we prefer automatic methods over manual coding because the latter are expensive and time consuming. Thus, we first try to use methods in prior literature [21, 22] that can automatically measure political bias of URLs by analyzing their sharing patterns on Twitter. Since these methods require a large number of tweets containing the news URLs, we cannot use them because very few of the personalized search results (URLs) are frequently tweeted. Therefore, we use mediabiasfactcheck.com which provides the political bias of 1,540 media sources (identified as domains) on a 5-point scale: left, left-center, center, right-center, and right [4]. For further quantification, we convert this 5-point scale to specific scores as: left = -100, left-center = -50, center = 0, right-center = 50, and right = 100. Note that mediabiasfactcheck.com provides the political bias for domains (e.g., nytimes.com) not URLs. Since we cannot automatically measure the political bias of personalized news URLs, we estimate their political bias using their domain's political bias from mediabiasfactcheck.com.

3 EVALUATION

Table 2 reports the personalization results quantified as pairwise difference and edit distance across pro-immigration (P), anti-immigration (A), and control (C) profiles.

In terms of difference, although it varies across different search terms, the average difference between pro-immigration and anti-immigration ($P - A$) is 3.2%, between pro-immigration and control ($P - C$) is 3.9%, and between anti-immigration and control ($A - C$) is 3.8%. Using the standard t-test, we are able to conclude that these average differences are significantly different from zero at 0.0001 significance level. The average edit distance between pro-immigration and anti-immigration is 8.4, between pro-immigration and control is 10.3, and between anti-immigration and control is 10.3. Using the standard t-test, we are again able to conclude that these edit distances are significantly different from zero at 0.0001 significance level. The edit distance values are higher than the difference because edit distance not only considers the difference in two lists of search results but also the changes in relative rankings of search results. Since difference and edit distance show the same trend in personalization across different search terms, we focus on the difference metric for discussion in the rest of this section.

To analyze personalization across different search terms, we sort search terms in Table 2 in descending order of $P - A$ values for each policy category. Note that the noise in search results due to A/B testing among four identical browser profiles for both pro- and anti-immigration is 0.5% on average. Thus, we focus our attention on search terms for which $P - A$ average difference exceeds 5% because it is at least an order of magnitude more than the average noise. A total of nine search terms meet this criterion: (1) comprehensive immigration reform, (2) uninsured Americans, (3) medicare for all, (4) national debt, (5) flat tax, (6) pro-life, (7) Paris climate agreement, (8) carbon footprint, and (9) support our veterans.

The comparison of pro-immigration versus anti-immigration and control profiles for these search terms reveals interesting insights.

Policies	Search Terms	Difference (%)			Edit Distance		
		$P - A$	$P - C$	$A - C$	$E(P, A)$	$E(P, C)$	$E(A, C)$
Immigra-tion	Comprehensive immigration reform	**6.3**	18.2	13.5	14.7	38.5	26.5
	White nationalism	4.3	4.2	2.7	8.5	8.7	7.2
	Dream Act	3.0	5.8	4.0	6.7	15.8	11.5
	Anchor babies	2.5	4.3	6.8	5.8	14.2	20.0
	Illegal immigrants	1.2	2.3	1.2	5.7	14.2	9.2
Foreign Policy	ISIS	1.7	0.8	1.8	4.3	2.3	5.3
	Benghazi	1.3	5.7	6.0	4.8	13	15.8
	Syria war	1.0	1.2	1.0	3.0	2.7	3.3
	Iran deal	0.5	0.5	0.3	0.8	1.0	0.5
	Aleppo	0.0	0.2	0.2	0.0	0.2	0.2
Health-care	Uninsured Americans	**7.2**	11.3	10.5	15.7	24.2	22.5
	Medicare for all	**6.2**	10.0	4.8	16.2	26.0	11.5
	Health insurance	4.3	2.0	3.0	15.3	7.2	9.5
	Affordable Care Act	2.7	5.0	6.8	8.2	17.8	24.5
	Obamacare	0.3	4.3	4.3	0.7	14.5	14.7
Economy	National debt	**12.0**	16.5	12.2	27.3	39.5	30.7
	Flat tax	**6.7**	6.2	1.0	13.0	12.0	2.0
	NAFTA	2.8	3.3	6.0	7.0	10.8	17.3
	Wall street	2.8	0.3	2.5	6.8	0.7	6.2
	Federal budget	1.5	1.3	0.7	10.2	8.2	5.0
Abortion	Pro-life	**5.7**	2.3	4.0	19.3	7.3	13.3
	Planned parenthood	3.5	7.0	5.3	9.0	14.3	13.0
	Roe v. Wade	2.8	2.2	4.3	7.0	5.8	11.8
	Pro-choice	1.7	3.0	4.0	4.8	6.7	10.7
	Women's rights	0.5	2.8	2.8	2.0	16.0	16.2
Gay Rights	LGBT	4.3	3.2	7.0	14.5	7.7	21.2
	Traditional marriage	4.2	2.8	2.8	9.7	6.2	6.0
	Gay couple	3.8	3.3	4.3	10.2	8.0	11.7
	Marriage equality	3.0	2.8	4.3	8.2	7.2	11.8
	Same-sex marriage	2.0	1.8	1.7	5.0	4.7	3.5
Gun Control	Gun license	4.3	2.2	5.7	10.8	5.0	14.3
	Background checks	4.3	2.2	2.8	13.8	6.7	8.2
	NRA	3.0	0.3	3.0	8.8	0.7	8.7
	Gun control	2.2	3.8	2.3	5.2	9.8	6.8
	Gun accessibility	0.7	0.5	0.2	5.8	3.2	2.7
Climate Change	Paris climate agreement	**7.2**	7.2	8.8	14.2	14.2	15.7
	Carbon footprint	**5.2**	5.3	3.0	11.8	12.3	5.8
	Climate debate	4.5	3.0	2.3	13.8	8.7	7.0
	Greenhouse gases	2.2	2.2	8.8	5.0	4.7	16.0
	Global warming	0.2	2.8	2.7	0.3	11.5	11.2
Educa-tion	No Child Left Behind	5.0	4.7	0.8	9.3	8.5	1.5
	Department of Education	3.3	2.7	1.5	12.2	9.7	6.8
	College affordability	2.8	3.0	0.2	10.3	10.2	0.7
	Race to the Top	1.0	1.8	2.3	3.8	6.5	9.2
	Free community college	0.3	2.7	2.8	0.7	6.0	6.3
Veterans	Support our veterans	**7.2**	5.8	5.5	15.5	13.2	12.7
	Veterans affairs	3.5	3.7	1.3	9.7	11	3.7
	Veterans	1.2	1.3	1.3	3.7	3.7	4.5
	Veteran benefits	0.8	1.2	1.2	1.7	3.0	3.0
	PTSD	0.5	3.3	3.2	0.8	8.7	8.2
Average		**3.2**	**3.9**	**3.8**	**8.4**	**10.3**	**10.3**

Table 2: Personalization (quantified using difference and edit distance) for 50 search terms. Note that $E(P, A)$, $E(P, C)$, and $E(A, C)$ are pairwise edit distance among pro-immigration, anti-immigration, and control profiles.

Top-k Search Results	Difference (%)			k-Edit Distance		
	$P-A$	$P-C$	$A-C$	$E(P, A)$	$E(P, C)$	$E(A, C)$
k = 10	2.3	2.4	2.7	5.7	5.7	6.7
k = 20	2.2	2.4	2.5	6.4	6.3	7.4
k = 30	2.4	2.5	2.7	6.6	7.3	8.0
k = 40	2.6	2.9	3.0	7.1	8.3	8.7
k = 50	2.7	3.2	3.1	7.5	8.9	9.1
k = 60	3.0	3.4	3.3	7.9	9.4	9.5
k = 70	3.2	3.7	3.6	8.3	9.9	9.9
k = 80	3.2	3.8	3.7	8.5	10.1	10.2
k = 90	3.3	3.8	3.7	8.6	10.2	10.2
k = 100	3.2	3.9	3.8	8.4	10.3	10.3

Table 3: Average personalization for all 50 search terms according to top-k (k∈{10,20,...,100}) ranked search results. We normalize edit distance (as k-edit distance) to make a fair comparison across different k values.

First, we observe that five of them (comprehensive immigration reform, uninsured Americans, national debt, Paris climate agreement, and support our veterans) also have high $P-C$ and $A-C$ differences. For example, national debt search term has 12.0% $P-A$ difference, 16.5% $P-C$ difference, and 12.2% $A-C$ difference. This shows that both pro- and anti-immigration profiles receive personalized search results that are different from each other as well as different from the control. Second, we observe that three of them (flat tax, medicare for all, and carbon footprint) have high $P-C$ difference but low $A-C$ difference. For example, flat tax search term has 6.7% $P-A$ difference, 6.2% $P-C$ difference, and 1.0% $A-C$ difference. This shows that the pro-immigration profile receives personalized search results that are different from both the control and anti-immigration profile. Third, we observe that only one of them (pro-life) has a high $A-C$ difference but low $P-C$ difference. Specifically, pro-life search term has 5.7% $P-A$ difference, 2.3% $P-C$ difference, and 4.0% $A-C$ difference. This shows that the anti-immigration profile receives personalized search results that are different from both the control and pro-immigration profile.

Note that there are search terms for which $P-A$ difference is not high (e.g., < 1.5%), but it does not necessarily mean that there is no personalization because their differences to control (both $P-C$ and $A-C$) may be high. This is because both pro- and anti-immigration profiles receive similar personalized search results with respect to the control so they are not much different from each other. For example, Benghazi search term has only 1.3% $P-A$ difference, but 5.7% $P-C$ difference and 6.0% $A-C$ difference.

Next, we study the rank of personalized search results to gauge whether they disproportionately appear at the top or bottom of the list of results. Table 3 reports the average personalization for all search terms based on the top-k (k∈{10,20,...,100}) search results. We note that average $P-A$ difference started at 2.3% for k=10, slightly increases for increasingly k values, and reaches 3.2% for k=100. We observe a similar trend for $P-C$ and $A-C$ differences as well as in terms of edit distance. Note that we normalize edit distance as k-edit distance (edit distance/k) for a fair comparison across different k values. Overall, while personalization slightly increases at bottom ranks, we conclude that personalization remains substantial for top ranked search results. Since personalization (in terms of both difference and edit distance) for several search terms exceeds the average by multiple factors, we gather that there exists significant personalization even in top-10 search results.

Search Terms	Political Bias		
	$B(P)$	$B(A)$	$B(P)-B(A)$
Carbon footprint	-30.0	10.0	-40.0
Comprehensive immigration reform	-35.3	-3.1	-32.2
Paris climate agreement	-23.3	-5.9	-17.4
Pro-life	36.1	52.8	-16.7
Support our veterans	14.3	25.0	-10.7
Uninsured Americans	-18.8	-10.9	-7.9
Flat tax	-4.5	-13.3	8.8
Medicare for all	-27.8	-37.0	9.2
National debt	11.5	-6.5	18.0
Average	-8.6	1.3	-9.9

Table 4: Political bias for nine top-personalized search terms. Note that $B(P)$, $B(A)$, and $B(P)-B(A)$ respectively are political bias of personalized search results for the pro-immigration and anti-immigration profiles, and the difference between their political bias. Table is sorted in ascending order of $B(P)-B(A)$ values. A negative $B(P)-B(A)$ value indicates the pro-immigration profile received more Democratic-leaning personalized search results, while a positive $B(P)-B(A)$ value indicates the pro-immigration profile received more Republican-leaning personalized search results.

We further analyze political bias of the personalized search results for the nine most personalized search terms in Table 2. Using mediabiasfactcheck.com, we are able to estimate the political bias of 373 news stories out of 792 (47%) personalized results for these nine most personalized search terms. Table 4 reports the average political bias of personalized search results for the pro- and anti-immigration profiles, and the difference between their political bias averages. The average political bias of pro-immigration profile is -8.6 which is Democratic-leaning and that of anti-immigration profile is 1.3 which is Republican-leaning (negative values represent Democratic-leaning and positive values represent Republican-leaning). In other words, the pro-immigration profile receives more Democratic-leaning personalized search results than that of the anti-immigration profile. We use the Kolmogorov-Smirnov test [25] to compare the political bias distributions of personalized search results for pro- and anti-immigration profiles. We are able to reject the null hypothesis that both distributions are the same at the 0.05 significance level. Thus, we conclude that the search terms receiving most personalization tend to get personalized results that reinforce the presumed partisanship. Note that political bias of personalized search results varies across different search terms. Personalization reinforces the presumed partisanship for six out of nine (carbon footprint, comprehensive immigration reform, Paris climate agreement, pro-life, support our veterans, and uninsured Americans) search terms that received most personalization. For the remaining search terms, it counters the presumed partisanship.

4 RELATED WORK

The most ubiquitous application of personalization is in online behavioral advertising, which targets users with personalized ads based on their browsing history. Several studies have reported personalization in online advertising [8, 11, 14, 41]. Researchers have also investigated personalization in terms of price steering and price discrimination on ecommerce sites [17, 27, 28, 40].

A long line of research has looked at web search personalization. Hannak et al. [16] examined several features such as browser user-agent, user's ethnicity, browsing history, and search query that Google can use to personalize search results. However, they only found significant personalization based on Google account login status and IP geolocation. In a follow-up [18], the author extended this study to Bing and DuckDuckGo. They found that, on average, 15.8% search results on Bing are different due to personalization based on the same features as on Google. They found no significant personalization on DuckDuckGo.

To further examine the role of location in triggering personalization, Xing et al. [42] used a browser extension called Bobble and reconfirmed that a user's location (inferred by geolocating IP addresses) is the most dominant trigger factor for personalization. Furthermore, Kliman-Silver et al. [20] examined the relationship between distance and location-based personalization, and the impact of location-based personalization on different types of queries. They found that the differences between search results grow as the physical distance between locations of the users increases.

Puschmann [32, 33] analyzed crowdsourced Google search results from approximately 4,000 users for 16 terms related to German political parties and candidates [3]. He found that search results are personalized primarily based on location, language, and time. He concluded that a small fraction of personalized search results, which could not be explained based on the aforementioned factors, may be trigged by latent (or non-observable) factors such as browsing history and Google account information that they did not collect from users.

Robertson et al. [35] audited personalization on Google Search by asking crowd workers to install a browser extension which collected Google search results for a variety of political search queries in standard and incognito modes. They found significant personalization based on Google account login status, self-reported usage of Google/Alphabet services such as YouTube, and self-reported rating of Trump but not self-reported political party affiliation. In a follow-up [34], the authors reported that personalized search results have minimal differences in terms of political bias.

While there is evidence of personalization in ads and ecommerce sites based on browsing history [8, 41], prior research has found Google search personalization primarily due to geolocation and account login status. In contrast to prior work, we believe that our sock puppet auditing system can help reveal the impact of browsing histories, which are explicitly designed to reflect different political ideologies, on search personalization.

5 DISCUSSION AND CONCLUSION

Overall, we observe significant personalization based solely on browsing history. The personalized results tend to reinforce the presumed partisanship of browser profiles that are trained by visiting web pages reflecting divergent political discourses. Our findings provide further empirical evidence for the underlying causes of filter bubbles or echo chambers. The explanation of why we found evidence of significant web search personalization while past work did not could be due to the following.

First, our research is different from prior work in terms of the methodology in training browser profiles. While previous work [15, 16] trained browser profiles to reflect different demographics, we trained browser profiles to reflect different political stances. Specifically, both Hannak et al. [16] and Haim et al. [15] trained browser profiles to represent different demographic groups such as gender, age, income, lifestyle, and ethnicity. In contrast, we trained browser profiles to explicitly reflect opposing views on the topic of immigration rather than different demographics. Specifically, we trained browser profiles using news stories posted by Twitter accounts who clearly demonstrated distinct political stances on the topic of immigration.

The second reason could be the difference in search terms that were used to test personalization. Prior literature reported that different search terms can trigger different magnitudes of personalization [16, 21]. Unlike previous work [15, 16, 35] which used search terms covering a variety of topics popular at the time, we used search terms related to the training topic of immigration. Specifically, Haim et al. [15] used general search terms such as "Germany". Hannak et. al. [16] used general search terms about topics such as news sources and literature. Robertson et al. [35] used general political search terms such as "US President." In contrast, after training browser profiles reflecting opposing views on the topic of immigration, we used immigration-related search terms such as "comprehensive immigration reform" and "illegal immigrants". We also used search terms about other relevant political topics [23] such as foreign policy.

Last but not least, the changing nature of Google's personalization algorithm could be another reason. Google is known to continuously tinker personalization algorithms as well as update their data sources over time. For example, Google changed its privacy policy in 2012 [2] and more recently in 2016 [6] allowing them to combine user data collected across all of its services (e.g., Search, Gmail, Google Analytics, DoubleClick) for targeted advertising and content personalization. Thus, Google can now more effectively personalize search results based on a user's browsing history inferred from its third-party analytics and tracking network. Since personalization algorithms are continuously being tweaked, we plan to longitudinally study personalization for different search terms over an extended period of time as part of our future work.

Prior work using controlled experiments as well as surveys of real users [15, 16, 32–35] has not found evidence of significant web search personalization based on browsing history. While simulating behaviors of real users completely will always be impossible, controlled experiments such as our sock puppet auditing system enable us to isolate the effect of different types of browsing history on search personalization. Thus, in this paper, we studied search personalization on Google News in a controlled setting by using browser profiles that were specifically trained to reflect strongly divergent opinions on the topic of immigration. Using this strategy, we find evidence of significant search personalization on Google News. This finding creates an opportunity to conduct further research on other factors that can trigger search personalization.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation (grant number 1715152) and by the Obermann Center for Advanced Studies.

REFERENCES

[1] Selenium webdriver. https://www.seleniumhq.org/projects/webdriver/.

[2] Google user data to be merged across all sites under contentious plan. The Guardian. http://bit.ly/2DluqbV, 2012.

[3] Data donation - German Parliamentary Election 2017. http://bit.ly/2t9QcLf, 2017.

[4] Media bias/fact check - the most comprehensive media bias resource. https://mediabiasfactcheck.com/, 2018.

[5] W. Andrews and T. Kaplan. Where the candidates stand on 2016's biggest issues. the new york times. http://nyti.ms/2ERzPEn, 2016.

[6] J. Angwin. Google Has Quietly Dropped Ban on Personally Identifiable Web Tracking. ProPublica. http://bit.ly/2eAjC9w, 2016.

[7] E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 2015.

[8] R. Balebako, P. Leon, R. Shay, B. Ur, Y. Wang, and L. Cranor. Measuring the effectiveness of privacy tools for limiting behavioral advertising. In *Web 2.0 Workshop on Security and Privacy (W2SP)*, 2012.

[9] P. R. Center. Political Polarization in the American Public: How Increasing Ideological Uniformity and Partisan Antipathy Affect Politics, Compromise and Everyday Life. http://pewrsr.ch/2n8oI2I, 2014.

[10] F. J. Damerau. A technique for computer detection and correction of spelling errors. *Communications of the ACM (CACM)*, 7:171–176, 1964.

[11] A. Datta, M. C. Tschantz, and A. Datta. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. In *Privacy Enhancing Technologies Symposium (PETS)*, 2015.

[12] S. Englehardt and A. Narayanan. Online tracking: A 1-million-site measurement and analysis. In *ACM Conference on Computer and Communications Security (CCS)*, 2016.

[13] M. Garcia. Racist in the Machine: The Disturbing Implications of Algorithmic Bias. *World Policy Journal*, 33(4):111–117, 2016.

[14] S. Guha, B. Cheng, and P. Francis. Challenges in measuring online advertising systems. In *ACM Internet Measurement Conference (IMC)*, 2010.

[15] M. Haim, A. Graefe, and H.-B. Brosius. Burst of the filter bubble? effects of personalization on the diversity of google news. *Digital Journalism*, 2017.

[16] A. Hannak, P. Sapiezynski, A. M. Kakhki, B. Krishnamurthy, D. Lazer, A. Mislove, and C. Wilson. Measuring personalization of web search. In *ACM International Conference on World Wide Web (WWW)*, 2013.

[17] A. Hannak, G. Soeller, D. Lazer, A. Mislove, and C. Wilson. Measuring price discrimination and steering on e-commerce web sites. In *ACM Internet Measurement Conference (IMC)*, 2014.

[18] A. Hannï£¡k, P. Sapie?y?ski, A. M. Khaki, D. Lazer, A. Mislove, and C. Wilson. Measuring personalization of web search. https://arxiv.org/abs/1706.05011, 2017.

[19] K. H. Jamieson and J. N. Cappella. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press, 2008.

[20] C. Kliman-Silver, A. Hannak, D. Lazer, C. Wilson, and A. Mislove. Location, location, location: The impact of geolocation on web search personalization. In *ACM Internet Measurement Conference (IMC)*, 2015.

[21] J. Kulshrestha, M. Eslami, J. Messias, M. B. Zafar, S. Ghosh, K. P. Gummadi, and K. Karahalios. Quantifying search bias: Investigating sources of bias for political searches in social media. In *The 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2017.

[22] H. Le, Z. Shafiq, and P. Srinivasan. Scalable News Slant Measurement Using Twitter. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2017.

[23] H. T. Le, G. Boynton, Y. Mejova, Z. Shafiq, and P. Srinivasan. Revisiting The American Voter on Twitter. In *ACM Conference on Human Factors in Computing Systems (CHI)*, 2017.

[24] T. Libert. Exposing the Invisible Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites. *International Journal of Communication*, 2015.

[25] F. J. Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46:68–78, 1951.

[26] D. B. Melican and T. L. Dixon. News on the net: Credibility, selective exposure, and racial prejudice. *Journal of Communication Research*, 35:151–168, 2008.

[27] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Detecting price and search discrimination on the internet. In *ACM Workshop on Hot Topics in Networks (HotNets)*, 2012.

[28] J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Crowd-assisted search for price discrimination in e-commerce: first results. In *ACM Conference on emerging Networking EXperiments and Technologies (CoNEXT)*, 2013.

[29] S. U. Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. 2018.

[30] A. Pansari and M. Mayer. This is a test. this is only a test. google official blog. http://bit.ly/2DngZbg, 2006.

[31] E. Pariser. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. The Penguin Group, 2012.

[32] C. Puschmann. How significant is algorithmic personalization in searches for political parties and candidates? http://bit.ly/2fEhDqK, 2017.

[33] C. Puschmann. How significant is algorithmic personalization in searches for political parties and candidates? (Part II). http://bit.ly/2E7n5KL, 2017.

[34] R. E. Robertson, S. Jiang, K. Joseph, L. Friedland, D. Lazer, and C. Wilson. Auditing partisan audience bias within google search. In *The 21st ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW'18)*, 2018.

[35] R. E. Robertson, D. Lazer, and C. Wilson. Auditing the personalization and composition of politically-related search engine results pages. In *The 29th International World Wide Web Conference Committee (IW3C2)*, 2018.

[36] C. Sandvig, K. Hamilton, K. Karahalios, and C. Langbort. Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. In *Data and Discrimination*, 2014.

[37] T. Speicher, M. Ali, G. Venkatadri, F. N. Ribeiro, G. Arvanitakis, F. Benevenuto, K. P. Gummadi, P. Loiseau, and A. Mislove. On the Potential for Discrimination in Online Targeted Advertising. In *Conference on Fairness, Accountability, and Transparency (FAT*)*, 2018.

[38] N. Stroud. Polarization and partisan selective exposure. *Journal of Communication*, (60):556–576, 2010.

[39] J. Swales. *Genre analysis: English in academic and research settings*. 1990.

[40] T. Vissers, N. Nikiforakis, N. Bielova, and W. Joosen. Crying wolf? on the price discrimination of online airline tickets. In *Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs)*, 2014.

[41] C. E. Wills and C. Tatar. Understanding what they do with what they know. In *ACM Workshop on Privacy in the Electronic Society (WPES)*, 2012.

[42] X. Xing, W. Meng, D. Doozan, N. Feamster, W. Lee, and A. C. Snoeren. Exposing inconsistent web search results with bobble. In *International Conference on Passive and Active Network Measurement (PAM)*, 2013.

[43] D. Zillmann and J. Bryant. *Selective Exposure to Communication*. Routledge, 2013.

[44] H. G. d. Zuniga, T. Correa, and S. Valenzuela. Selective exposure to cable news and immigration in the u.s.: The relationship between fox news, cnn, and attitudes toward mexican immigrants. *Journal of Broadcasting and Electronic Media*, 2012.