Challenging Google Search filter bubbles in social and political information: Disconforming evidence from a digital methods case study

Cédric Courtois*, Laura Slechten, Lennert Coenen

School for Mass Communication Research, KU Leuven, Belgium

ARTICLE INFO

ABSTRACT

This article engages in the debate on supposed online 'filter bubbles' by analysing a panel of Google users' search results on a standardized set of socio-politically themed search queries. In general, the query results appear to be dominated by mainstream media sources, followed at a large distance by civil society and government resources. By means of mixed model regression analyses, with the prominence of different source types in the search results as dependent variables, it was tested whether search results vary across Google Search users. The results indicate that the inclusion of participants as a random effect does not explain variance when controlling for the different query keywords and the time at which the queries were ran. Hence, this study does not support the occurrence of 'filter bubbles' in Google Search results in the context of social and political information.

1. Introduction

The overwhelming availability of online information has pushed web developers into designing mechanisms that are able to cope with adverse consequences such as information overload and information anxiety (Bawden and Robinson, 2009). Search engines and especially filtering algorithms work to relieve users from these problems by selecting and prioritizing information into personally tailored selections of relevant information (Bozdag, 2013). Filtering and recommendation systems specifically aim to improve the ease of using platforms by suggesting appropriate and relevant information to individual users (Knijnenburg et al., 2012). This personalization process is usually based on the automated analysis of a broad array of personal data, which allows inferring user preferences, contexts, and interests. Thus, the algorithms exercise power in deciding the types of information that are prioritized over others, whether information is classified favourably, how it is associated with other kinds of information, and how it is eventually filtered and presented to users (Diakopoulos, 2015). It has been argued that through these characteristics, algorithms assume the qualities of a social institution (Napoli, 2014). More specifically, they are characterized by regulative, normative, and cultural-cognitive components that enable and constrain the flow of information (Christin, 2016). As such, they play a pivotal role in the construction of reality as tailors of frames of reference for their users (Just and Latzer, 2016). However, users of search algorithms are generally not transparently informed on how their data are processed and are urged to keep sharing information in order to provide an optimal user experience (Peacock, 2014). Due to the algorithms' proprietary nature, users are mostly unaware of how algorithms work, causing them to treat these mechanisms as unproblematic means to and end (Gillespie, 2014).

---

* Corresponding author.
  E-mail addresses: cedric.courtois@kuleuven.be (C. Courtois), laura.slechten@kuleuven.be (L. Slechten), lennert.coenen@kuleuven.be (L. Coenen).

In light of these considerations, the potential consequences of covert online information filtering has become the subject of concern and fierce debate, with some claiming that it is an indispensable factor in creating bubbles that narrow the scope of information web users come across (Pariser, 2011). It is argued that such constraints have a profound impact by guiding users' belief structures and the actions that stem from it. Currently, there is little rigorous empirical support for any of these claims, which is partly due to the complex nature of studying these phenomena (Ørmen, 2016). This study therefore aims to provide an empirical basis for the debate by systematically documenting and analysing the Google Search results of a panel of users, assessing (a) what information is prioritized, (b) whether individual differences in search results exist, and (c) whether such personalization occurs along the lines of users' ideological positions. The choice for Google is incited by the prominence of online search in acquiring novel knowledge and verifying prior information. It is important to emphasize that the focus of this study is primarily on the overall types of results, and the potential individual variance in Google Search results that is provoked by standardized search queries. Of course, in a naturalistic setting, Google users compose and refine their own queries, while assessing the results. However, to shed light on how Google treats search queries, a strictly controlled environment is a necessity. Still, rather than relying on dummy accounts, the standardized queries are run through actual users' accounts, as if they would search for them themselves. In the following sections, we first address the conceptual background and history of personalized search, as well as its potential vulnerabilities and consequences in order to set the scene for the empirical study.

1.1. Online search environments

Early online search engines analysed web content based on keywords and metatags (Seymour et al., 2011). Such engines faced various challenges, including coping with the massive growth and scale of online information, its dynamic and self-organising character, and the complex hyperlinked structure of the web. These inherent features of the web complicated early search engines' scalability and made them particularly vulnerable for spamming efforts that artificially boost the rankings of third party websites (Langville and Meyer, 2006). This situation changed profoundly by the end of the 1990s with the development of link analysis systems, of which Google Search is undoubtedly the most refined and successful business application up to date. More specifically, Brin and Page (2012) developed the PageRank algorithm at the University of Stanford within the context of a publicly-privately funded research project. PageRank's strength lies in appropriating the networked structure of the web. Its mathematics is based upon the straightforward principle that *'a webpage is important if it is pointed to by other important pages'* (Langville and Meyer, 2006, p. 26). Beyond this principle, the algorithm has since then been incessantly refined and supplemented by others, increasing its performance and resilience to spamming and large-scale manipulation (Bar-Ilan, 2007; Google, 2016).

These continuous developments occur within the context of a rapidly expanding media and technology conglomerate. What initially started as a non-commercial research project gradually developed into a promising commercial venture (Steiber and Alänge, 2013). Since its inception, Google has grown from a search engine provider into the diversified publicly traded enterprise Alphabet, catering the needs of over a billion users through a proliferation of adjacent services (Vaidhyanathan, 2012). Through this process, the development of a viable business model became an evident necessity (Finkle, 2012). Google now essentially harnesses refined user data from its various popular services to target the online advertising market. It lucratively mediates between advertisers and web platforms by enabling the former to individually target search results to specific consumer profiles, while also providing advertising revenues to the latter (Kang and McAllister, 2011; Lee, 2011). Consequently, the company's objective has become to invest in a wide array of integrated services that that keep the so-called 'virtuous cycle of big data' running (Harrison, 2015). This cycle implies that Google appropriates user data to continually improve services and its supporting algorithms to create a better experience. This, in turn, attracts and retains more marketable users, which increases the volume and the granularity of collected user data that supports further improvements and commercial micro-targeting efforts. Thus, the structural boundaries of the algorithms' design are inevitably informed by a clear set of values that are inevitably market-oriented rather than inspired by fairness and representativeness of information (Van Couvering, 2007).

It is estimated that Google annually changes its search algorithm about 500 to 600 times, which usually comprises minor adjustments (Moz, 2016). In 2009, however, Google introduced a pivotal novelty that formed the basis for search engine personalization (Horling, 2009). To enhance the search engine user experience, mechanisms were implemented to infer user preferences. This system is based on the intuitive logic that the relevance of information varies between users and their contexts. Hence, it infers and accounts for the variable meaning of the same keywords for different users. Divergent types of information sources were programmed into the algorithm: on the short term, recent search queries inform about context, whereas over time refined user profiles are built to get a grasp on long-term patterns of individual preferences and characteristics (Smyth et al., 2011).

Research on the occurrence of personalization in Google products has rendered mixed results. One of the first studies that focused on Google Search involved running a standardized set of 80 queries on actual user accounts, comparing its results with the results of the same queries run without logging in (Hannak et al., 2013). The results were than analysed for overlap (i.e., Jaccard Index) and differing rank orders (i.e. Kendall's Tau). The results indicated that, on average, about twelve per cent of searches were personalized, peaking for queries on news and political issues. A follow-up study, centred on the factor of location, showed that up to 34% of search results varied in terms of the results that were shown (Kliman-Silver et al., 2015). Apart from that, also substantial variations in rank orders of the search results were found. Nevertheless, personalisation does not seem to emerge in a straightforward manner. Hoang et al. (2015) experimented with training Google Search accounts, rendering differences in up to one quarter of the returned results. However, the outcomes of the training did not appear to directly correspond with the contents of the account training. Still, this evidence is somewhat tentative as it dates back several years, while the service further developed. Moreover, the majority of studies were based on analysing mock-up accounts. On the other hand, a variety of studies, particularly focused on Google News repeatedly

showed the absence of any meaningful personalization, apart from explicit suggestions (Cozza et al., 2016; Datta et al., 2015; Haim et al., 2018). This raises considerable doubt on whether personalization is observable in more recent versions of the Google Search environment.

1.2. Potential consequences of personalized search

Google's search engine personalization relies on learning users' preferences and the context of searches. Each search query adds to the formation of refined user profiles, in turn influencing future search results. Hence, there is a feedback loop running between users and the online search technology. This implies that certain information takes primacy over other information because it is filtered out or pushed down in the result ranks. Hillis et al. (2012, p. 15) argue that *'a virtuous circle of cybernetic feedback loops ensues. Search algorithms "learn" about our preferences and desires as they endlessly concatenate information about personal quests of individual users'*, which feeds a recursive system of *'adaptation and modulation'*. Pariser (2011) labelled such occurrence with the concept of the *filter bubble*, which resonates considerably within both academic and public debate. He warns for the potential risks of increasing online personalization because algorithmically produced recommendations bear the potential to narrow the scope of accessible online information: over time, users get to see more of the same, pushing users into the psychological comfort zone of self-confirmation and risking polarization on a societal level. The most pertinent danger, he continues, lies within the individualized, involuntary and invisible nature of such bubbles.

The opaqueness of how search results are compiled is perhaps the most troublesome, especially when the search and personalization algorithms are biased or manipulated to begin with. It has become clear that algorithms are all but neutral devices: they have in-built values and are prone to bias, similar to human actors. The reliance on machine learning to alter algorithms based on user data may even obscure platform developers' understanding of algorithmic interactions. In addition, third parties try to get their information visible and marked with higher priority through a wide range of practices that boost traffic to get higher up in search rankings (Gillespie, 2017). This raises concern as web users are known to trust Google Search results (Pan et al., 2007), and search engines in general, to neutrally filter and rank information according to its relevance (Penna and Quaresma, 2015). Research suggests that the majority of users favour the highest ranked results (Kammerer and Gerjets, 2014). This might be problematic because at the same time sponsored results have increased in prominence (Höchstötter and Lewandowski, 2009). This has led to considerable concern on the potential effects of search results and their rankings. A set of recent studies by Epstein and Robertson (2015) shows that the manipulation of rankings of a mock-up search engine, favouring one political candidate over others, has the potential to affect undecided voters' attitudes towards candidates as well as their voting intentions. Strikingly, even participants' awareness of ranking bias did not nullify the observed effects.

2. The present study

The present research is guided by three research questions regarding the functioning of filtering algorithms, focusing on Google Search. With an estimated global market share of 73 per cent, Google Search accounts for the vast majority of web searches (Netmarketshare, 2018). With such a dominant position in selecting and filtering information, it is hardly surprising that current debates on misinformation also involve its mechanisms in the supposed spreading of 'fake news' (Bakir and McStay, 2017; Brennen, 2017). To counter negative exposure within these debates, Google announced to widen the set of tools that allow users to report misleading or offensive content and suggestions (i.e. direct feedback tools), whereas the search algorithm itself was tweaked to demote low quality search results in the rankings (Gomes, 2017). In light of this debate, we first question what sources Google generally prioritizes in response to socio-political queries (RQ1).

Second, we assume that Google personalizes search results based on prior search and browsing activities, eventually tailoring itself around user preferences. However, as argued, prior research tends to diverge in methods, moments of data collection, and its resultant findings. Despite these existing studies, systematic research on the assertion of personalization that involves real users and their accounts remains scarce. We therefore aim to specifically analyse whether and to what extent users' results of Google Search queries vary across individual users (RQ2).

Tied to this second question we can ask what drives the process of personalization. Elaborating on the aforementioned logic of the cybernetic feedback loop, it would make sense that search personalization over time favours information that is similar to previously accessed information. This would imply that future search results tend to echo prior ones and form a bubble that accords to users' ideological positions. When this echo amplifies ideological positions, either liberal or conservative, it could be a feeding ground for increased polarization. Hence, assuming that search engine personalization exists, we ask whether this personalization is in accordance with users' prior search behaviour and ideological positions (RQ3).

To answer these questions, this study combines digital methods and traditional survey research. This is accomplished by involving a panel of actual users in a research design that performs standardized Google Searches and stores their results for further analysis. This information is subsequently merged with user self-reports through an associated online survey. In line with recommendations by Ørmen (2016), we will maximize internal validity by constraining several factors that affect search results. More specifically, search queries are controlled by design, which ties the explanation of the findings to variation across participants' search environments. This inevitably comes at the cost of limiting the external and ecological validity of the design. We fully acknowledge that in a naturalistic setting, users devise their own keywords and browse and compare results. However, in order to assess algorithmically-induced variation in results, keywords must be standardized to maintain a ground for comparison. The goal of this study is not to infer how users interact with search engines, how they evaluate information, or how they are possibly affected by it. It does focus on how

Google as a search engine handles queries while considering users' actual search environments. The resulting insights indicate how search algorithms treat queries differently across users, which is important to inform and assess results that stem from studies that do occur in naturalistic settings in which users devise their own queries. If we find no meaningful personalization, it strengthens the validity of such research, shielding it from criticisms of not accounting for personalization. If, on the contrary, we do find traces of substantial personalization, it bears the consequence that users' search environments differ beyond the studied interaction with the search engine. This would mean that personalization is a confounding factor that affects the setting and hence the outcomes of such studies.

3. Method

3.1. Sample and data collection

This study was conducted in April–May of 2017 in Flanders, Belgium's Northern, Dutch-speaking region. Belgium is a federal country, composed of three regions. Both federal and regional levels are governed by a democratically elected parliament. The media landscape involves a relatively strong public service broadcaster, providing television, radio, and online services. Both the commercial audiovisual and print markets are characterized by a fair number of channels and titles, albeit concentrated within concentrated market in which two dominant players are present (VRM, 2016).

The present study relies on a combination of digital methods and users' self-reports through an online questionnaire. Extensively trained student-researches visited participants at home, strictly following a formal research protocol. This led to a sample of 350 Dutch-speaking Flemings. The sample was built by these students, drawing on their own extended social networks. The sample construction was guided by quota that considered education level, age and gender. This led to a sample that varied in education level (42% obtained a secondary education degree at most, 58% was higher educated), age ($M = 40.79$ years old, $SD = 13.71$) and gender (52% male, 48% female).

Upon active informed consent, the participants were asked to log in with their Google accounts to a dedicated, local application installed on the researcher's personal laptop that was logged on to the participant's private local area network. The script ran through a clean headless browser. Once successfully logged in, a set of 27 standardized socio-political keywords were ran. These keywords were randomized prior to the experiment. The search queries were inspired by Flemish political party programmes and important topics of public debate at the time of the study. The search queries cover nine socio-political themes: (a) environment, (b) education, (c) economy, (d) healthcare, (e) migration, (f) national security, (g) poverty, (h) social security and (i) taxation. Each theme is covered by three neutrally phrased query keywords. Translations of these queries are listed in Table 1.

The first ten organic Google Search results were parsed by the script and written into a central research database. The search

Table 1
The standardized keywords used in this study, derived from political party programmes and public debates during the data collection. The keywords are translated from Dutch. The original Dutch phrasing is available upon request.

Socio-political themes	Keywords
Education	Quality education Flanders
	PISA Flanders education
	Equal opportunities Flemish education
Climate	Program climate Flanders
	Improve air quality Flanders
	Achieve climate goals Belgium
Poverty	Debate on poverty statistics
	Migration poverty
	Impact measures poverty
Migration	Influx of asylum seekers in Flanders
	Contribution refugees to society
	Migration cost society
National Security	State of emergency Belgium
	Increase security Flanders
	Measures against terrorism
Social Security	Affordability pensions
	Affordable elderly care
	Limit unemployment benefits in time
International Economy	Impact global economy Belgium
	Wage cost disparity
	Multinationals investments Belgium
Taxes	Tax pressure Flanders
	Consequences tax shift
	Notional interest consequences
Healthcare	Privatizing healthcare
	Activate long-term sick people
	Savings healthcare Belgium
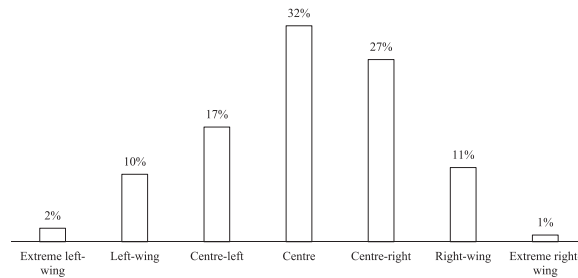
Fig. 1. Histogram of participants' ideological position.

results included a URL, a title, a description and its rank on the result page. Occasionally, the script was unable to perform all 27 queries. On other occasions, not all queries returned exactly ten search results. About 4% of the anticipated search results were not captured. In total, 90,810 search results were retained. Each search result is nested within one of the 27 keywords and 350 participants that were involved in the study.

3.2. Self-reports: online survey

Besides harvesting search results, the participants were asked to complete an online questionnaire that captured general socio-demographic information such as age, gender, and education. Moreover, participants were asked to indicate their ideological position on the political spectrum on a seven-point Likert scale, ranging from left-wing to right-wing ($M = 4.12$, $SD = 1.26$; Fig. 1). They were also probed for the frequency to which they search political information on Google ('Never' – 23%, 'Less than monthly' – 45%, 'At least Monthly' – 26%, 'At least Daily' – 5%). Finally, a question verified whether participants' Google accounts were more than a month in use at the time of the study (73%) – 65% had their account for one year or even longer.

3.3. Content analysis: coding search results

In order to assess the nature of the search results, a list of 300 unique domain names were distilled from the results database (e.g., Wikipedia.org). This list of domains was manually coded, following a coding scheme that consists of sixteen exclusive categories that reflect 'source type'. The domains were coded to establish meaningful categories of sources in the search results. Drawing upon precise page URLs to determine personalisation most likely artificially inflates differences, what we deliberately tried to avoid. For instance, getting different articles from the same, or a very similar online newspaper would count as personalization. However, in our view, such differences do not bear much meaning in terms of problematic algorithmic personalization.

The coded sources types are: (a) local government, (b) regional government, (c) national, federal government, (d) international government – e.g., European Union, (e) mainstream media, (f) online-only media, (g) left-wing politics, (h) right-wing politics, (i) centre politics, (j) civil society organizations, (k) political blogs and forums, (l) opinion leaders and think tanks, (m) commercial entities, (n) academic institutions, (o) reference materials, and (p) supervisory bodies.

Two researchers independently coded the 300 domain names. This led to an initial Krippendorff's α of .76. Initially inconsistently coded domains were subsequently coded by a third researcher and then discussed until a definitive consensus on a single value was reached for all domains.

3.4. Final data preparation: unit of analysis

Prior to the analysis, all three data sources were merged into a single dataset. The harvested raw search results ($N = 90,810$) were enriched with the content coding variables ($N = 300$) and the self-report measures ($N = 350$). Next, the content coding variables were multiplied with the inverse result rank (i.e. 11 – rank). This manipulation is required to address more weight to high ranking results than to lower ranking results. As mentioned, in an ecological setting, search engine users are inclined to address more weight and credence to high-ranking results than to low-ranking results. Rather than artificially keeping presence and ranking apart (i.e. by separately calculating Jaccard Index for occurrence and Kendall's Tau for rank order), we explicitly argue in favour of combining this information into a single measure.

Subsequently, the results were condensed into a single record per search query, per participant. This implies that the individual search result coding variables were summed into a single value per coding variable per query, per person ($N = 9142$). In principle, the coding variables can take on values up to 55 (i.e., summing all the possible ranks for the first ten search results). Hence, it should be emphasized that the unit of analysis in this study are the aggregated, weighted results of a search query ran from a particular Google account.

4. Results

The first research question focuses on the source types that are prioritized by Google. Table 2 summarizes the types of sources

Table 2
Relative presence of source types per search query category. Raw percentages (%r) are not weighted for result rank, whereas the weighted percentages are (%w).

	Education		Climate		Poverty		Migration		National Security		Social Security		International Economy		Taxes		Healthcare		Sample Mean	
	%r	%w	%r	%w	%r	%w	%r	%w	%r	%w	%r	%w	%r	%w	%r	%w	%r	%w	%r	%w
Local Government	39	51	40	47	2	2			1	1					3	2			1	1
Regional Government	3	3	10	11			4	3	4	2			4	6	4	6			11	13
National Government					12	12			7	8	3	5	10	14			3	2	5	6
International Government									2	2										
Mainstream Media	26	24	21	19	32	35	31	38	62	73	47	50	28	31	48	54	49	55	38	42
Online-Only Media			1	0	4	1	10	7	7	6	13	13	2	1	10	7	4	7	6	5
Left-Wing Politics			1	1							6	5	2	1	1	1	10	6	2	1
Right-Wing Politics							6	8	4	4	8	4			4	2			3	2
Center Politics							1	0					1	0					0	0
Civil Society	17	11	13	14	27	27	24	25	2	1	8	5	10	7	10	8	25	23	15	13
Political Blogs and Forums	1	0	1	0	1	0	5	5	1	0									1	1
Opinion Leaders and Think Tanks	1	1	1	1	7	9	3	4											1	2
Commercial Entities	8	2	5	3	9	8	8	5	7	1	3	4	20	16	11	11	3	4	7	5
Academic Institutions	4	7	4	2	4	3					10	13	15	15	3	3			6	6
Reference Materials (Wikipedia and Google Books)	2	1					7	5	2	1					4	6	4	3	2	2
Supervisory Bodies			2	1									9	9					2	2

present in the results to the nine query categories, both in raw and weighted percentages. The former percentage scores are un-weighted, and account only for mere presence without correcting for rank order. The latter do correct for the rank order in which results were displayed. The sample mean indicates that 38% of query results referred to mainstream media sources. This rises to 42% when the result rankings are taken into account. Moreover, mainstream media sources are represented throughout all socio-political themes, whereas only-online professional media outlets are far less present. In second instance, civil society organizations account for 15% of the raw results, which is however toned down to 13% when accounting for rank. Similar to mainstream media, this source type also covers all themes, albeit showing particular presence in light of poverty, migration, and healthcare issues. This is not the case for regional government sources, accounting for 11% of the raw, and 13% of the weighted search results. Interestingly, political parties are hardly present in the search results, regardless of their ideological positions. Still, when they are shown, their presence aligns with the themes they particularly focus on (e.g., migration for right-wing parties, healthcare for left-wing parties). Similarly, reference materials, including Wikipedia and Google Books are hardly present. The many blanks in Table 2 clearly show that a considerable number of sources are not considered. The obvious differences in, and even absence of particular sources leads to the conclusion that the type of search query has a considerable impact on the kinds of source types that are covered.

The second research question followed from conceptually problematizing the potential occurrence of personalization. So far, it is clear that different search queries lead to strongly differing search results. The question now is whether there is additional variance in the results' source types that is explained by individual participants. Moreover, as the study ran for nearly two months, it is not unlikely that the day on which queries were ran induced additional variance. To decompose these sources of variance, and ultimately test whether there is variance on the personal level, we turn to (generalized) linear mixed modelling. Because a search result is nested in a keyword that is run on a particular day, from a particular user account, multiple sources of variation are possible. All of these possibilities need to be considered and tested.

Further issues are the nature and distribution of the source type variables. The source type variables are in essence count vari-ables, weighted for rank. Moreover, as shown in Table 3, the majority of source type variables have a mean that is close to or even substantially larger than its standard deviation. For that reason, we opt for Poisson regression, albeit accounting for apparent overdispersion, i.e. excessive variance. This is a problem because Poisson regression assumes that mean and variance are similar to each other, which is not the case for these data. However, drawing on evidence from simulation studies, an additional observational-level random effect, besides those of queries and participants, is included to control for this disturbance (Harrison, 2014). In one case, we chose to test against a standard Gaussian distribution. A visual inspection of the mainstream media source type variable showed a fairly normal distribution, while its mean is substantially larger than its standard deviation. Also, while performing the analysis, the Q-Q plots of the residuals indicated no substantial deviation from a normal distribution. The analyses were performed in R, using the lme4 package (Table 4).

To break down and compare the potential origins of the variance in the source type variables weighted for rank, three types of null models are computed and then compared against each other. The first type of model consists of two random effects (Model A): the overdispersion correction that involves a random effect with a level for each observation and a random effect that reflects the different keywords tested in the study. Subsequently, this model is extended by incorporating a random effect that considers the timing of when the search query was ran. This grouping variable indicates the day on which the queries were ran on a participant account (Model B). Finally, a third model involves yet another random effect. This additional random effect groups participants (Model C). Hence, this final model tests whether including participant variation renders a better-fitting model. A χ^2-test is used to determine whether a subsequent model performs better than a prior one.

Even a visual analysis of the association between query categories and source types in the search results shows substantial differences that exceed chance by far. The formal analysis of adding grouping variables reveals that factoring in time as a random factor significantly improves model fit. This means that there was substantial variation in search results during the study, even when controlling for the differing search queries. However, adding participants as an additional random effect does not yield any better-

Table 3
Descriptive statistics of the search query results per source type.

	N	*M*	*SD*	Skew	*SE* Skew	Kurtosis	*SE* Kurtosis	Range
Local Government	1382	1.91	2.21	.87	.07	−.35	.13	9
Regional Government	4154	15.66	15.21	.96	.04	−.12	.08	55
National Government	3806	7.99	7.02	.86	.04	.15	.08	30
International Government	698	1.57	2.31	1.25	.09	.61	.19	9
Mainstream Media	9142	22.74	13.97	.11	.03	−.96	.05	54
Online-Only Media	8445	2.80	3.80	1.34	.03	1.12	.05	21
Left-Wing Politics	3131	2.36	2.93	.98	.04	−.12	.09	13
Right-Wing Politics	3257	3.15	4.51	2.05	.04	3.82	.09	24
Center Politics	818	.43	.88	3.07	.09	13.25	.17	8
Civil Society	8096	8.19	7.33	1.00	.03	.34	.05	35
Political Blogs and Forums	1383	2.25	3.32	1.27	.07	.09	.13	11
Opinion Leaders and Think Tanks	1729	4.35	5.00	.90	.06	−.35	.12	23
Commercial Entities	4987	4.64	5.15	1.00	.04	−.22	.07	26
Academic Institutions	4500	6.71	5.34	.50	.04	−.77	.07	21
Reference Materials	2444	3.07	3.39	1.01	.05	−.13	.10	11
Supervisory Bodies	2411	3.57	3.77	.79	.05	−.45	.10	15

ARTICLE IN PRESS

C. Courtois et al.   Telematics and Informatics xxx (xxxx) xxx–xxx

Table 4

Overview of null model tests (i.e., random intercepts-only models). All models are tested against a Poisson distribution, expect for the model on Mainstream Media (°), which is tested against a Gaussian distribution and does not contain an observation-level random effect.

	Null Model A Observation + Query		Null Model B Model A + Time		Null Model C Model B + Participant		Comparison A-B		Comparison B-C	
	BIC	LL	BIC	LL	BIC	LL	χ^2	p	χ^2	p
Local Government	4287	−2133	4131	−2051	4139	−2051	162.63	***	.00	
Regional Government	18,071	−9023	18,070	−9018	18,078	−9018	97671.00	***	.00	
National Government	17,016	−8496	16,960	−8463	16,968	−8463	64902.00	***	.00	
International Government	1986	−980	1986	−980	1992	−980	.00		.00	
Mainstream Media°	49,993	−24983	49,863	−24913	49,872	−24913	139.06	***	.00	
Online-Only Media	21,897	−10935	21,906	−10935	21,915	−10935	.537		.00	
Left-Wing Politics	10,288	−5132	10,074	−5021	10,083	−5021	221.91	***	.00	
Right-Wing Politics	11,676	−5826	11,642	−5805	11,650	−5805	41.83	***	.00	
Center Politics	1383	−681	1252	−613	1259	−613	137.30	***	.00	
Civil Society Organizations	38,516	−19244	38,483	−19224	38,492	−19224	41513.00	***	.00	
Political Blogs and Forums	3610	−1794	3498	−1735	3506	−1735	118.42	***	.00	
Opinion Leaders and Think Tanks	622	−3099	5688	−2829	5696	−2829	539.30	***	.00	
Commercial entities	19,620	−9798	19,207	−9586	19,215	−9586	422.35	***	.00	
Academic entities	19,749	−9862	19,733	−9850	19,742	−9850	24.19	***	.00	
Reference Materials	7589	−3783	7597	−3783	7605	−3783	.23		.00	
Supervisory Bodies	8932	−4454	8770	−4369	−4369	−4369	169.55	***	.00	

*** $p < .001$.

fitting models. Consequently, we need to conclude that variation in the results is not due to particular individual differences. We find no evidence for personalization within the collected data.

This means that the third research question, inquiring what personal factors might explain personalization becomes redundant. If there is no variation to be found on the individual level it does not make sense to include explanatory variables on that level, such as age, gender, and ideological position.

5. Discussion

This study subscribed to current debates on the algorithmic filtering and personalization of search results. Google has been particularly at the centre of these discussions, considering its market leadership in the Western world. The possibility of filter bubbles has found considerable traction both in popular and academic debate, warning for the potentially polarizing effects of algorithmically-governed information systems. Unfortunately, with regards to the Google Search case, this argument has been rarely supported by systematic research on a substantial set of real user accounts. With this study, we attempted to add to this debate by compiling a dataset based on a diverse user panel: more than three-quarters of the 350 panel participants had their Google account for over a year, whereas one third claimed to search for political information at least monthly. Moreover, participants varied considerably in their ideological positions.

In response to the first research question, the overall distribution of the search results tends to favour the primary and secondary definers that for long have been dominating the public debate. First and foremost, online versions of traditional legacy media, such as newspapers and broadcasters, relatively account for most of the search results. The queries only led to civil society and government sources in second and third instance. In general, there appears to be very little room for smaller entrants, such as novel journalistic initiatives that focus on online activities. Similarly, political parties' communications are hardly present in the collected data. The bypassing of traditional media outlets presumably takes place more often on social media platforms. Interestingly, reference materials, including Wikipedia and Google Books are hardly present, whereas previous research showed that Wikipedia is generally favoured as a source by Google (Silverwood-Cope, 2012). This finding seemingly contradicts prior research, although it is likely explained by the fact that the query keywords point into the direction of particular public debates and consist of deliberate combinations of words that are generally not covered by specific entries in the Dutch-language Wikipedia. Finally, sources that could be labelled as blatant misinformation, focusing on conspiracy theories and lacking sources and alternative views, were not encountered while coding the search results.

With regards to the personalization question, the current data do not show any indications that users in our panel received markedly different results that are likely attributable to their prior behaviour and preferences. The variation that we did find, apart from that caused by differing search queries, was explained by factoring in time rather than individual differences. On the one hand, this makes sense in a way that current affairs likely affect what sources are considered most relevant by the search algorithm. On the other hand, this finding highlights the issue that searching for information with the same keywords, even in a relatively short time span, leads to diverging sources. Still, it should be emphasized that exposure to political information is of course not dominated by online search. As such, we concur with Dutton and Reisdorf (2017) who stress that exposure to online information is not solely determined by search. They demonstrate that internet users tend to rely on a multitude of sources that confront them with a variety of viewpoints.

In conclusion, this study does not support claims on the assumed filter bubbles in online search, whereas some previous studies did point into that direction. Still we need to emphasize that research on algorithmic information filtering still faces a multitude of challenges, both conceptually and methodologically. There are marked differences in conceptualizing personalization. The key question is whether it makes sense to label the mere occurrence of different URLs and or rank orders as a sufficient condition to talk about personalization. We argued that the criterion should be the type of information, rather than a particular outlet. Two online, perhaps even related newspapers can tell a very similar story, yet they would be classified as a different result. Still, their differences are likely trivial, at least in a media system that draws on internal pluralism. In a highly competitive media environment based on the ideal of external pluralism, we are faced with a different story. A reason why colleagues might find traces of personalization is potentially a byproduct of a rather open stance towards what personalization entails. In that respect, we adopted a restrictive view by coding sources in categories, which attenuates what otherwise could be considered as a difference. We believe that such a conservative interpretation is necessary, especially when taking into account the far-reaching assumptions that are generally related to information personalization, such as affecting attitudes and behaviours in the public.

Furthermore, we drew upon mixed model analysis. This allowed to decompose the variance sources tied to keywords, time, and users. This approach proved a fruitful, yet stringent way of analysing the data at hand. A constant issue with researching personalization is the danger of misattributing seeming differences to personalization because alternative explanations are overlooked and excluded. For instance, in this study, ignoring the time factor would lead to the erroneous conclusion that there is variance on the individual level. Yet, other, predominantly technical artefacts are equally present, such as for instance the inclusion of serendipity in algorithmic systems, as well as queries being handled at different data centres (Hannak et al., 2013). Also, it must be taken into account that algorithms are ever-changing, which makes it tedious to compare results that are especially far apart in time and even draw on varying data collection methods.

Still, this study is not a definite argument that personalization in online search is trivial or even non-existent. There are multiple reasons for that. First of all, this is a case study in a particular region. As the search results are dominated by mainstream media, which are internally pluralistic, it made sense to code them as a single category. This makes far less sense in a context in which partisan media are present. In a similar vein, Belgium's political landscape is characterized by a fair amount of plurality, in which almost twenty political parties have representatives in at least one of the countries parliaments. In a bipartisan political situation, polarization is more feasible and perhaps reflected in search results.

Finally, this study particularly focuses on social and political issues in organic search results. From a business perspective, there is little for Google to gain by allowing bias with regards to these issues. On the contrary, it leaves ample room for bad public relations, undermining the public's trust in its services. From that point of view, divergence in query results should be seen as collateral damage rather than a deliberate goal. This is not the case for information that is especially prone to commercialization and is intensively advertised through sponsored results that by default favour the highest bidder. In this particular case, such sponsored results did not appear during the data collection. Still the question remains whether political communication will widely engage in this kind of advertising in the near future. That is why the topic of potentially personalized online information stays relevant, despite the seeming current exaggeration in public debate. Still, for now, the evidence tends to stack against all too generalizing claims (see also e.g., Borgesius et al., 2016). We suspect that the fierceness of this debate is largely because of the enigmatic character of algorithms. That is why regardless of the available sobering evidence, the call for algorithmic transparency remains a valid one. It would aid researchers in their understanding of the role of algorithms in society, including those responsible for searching online.

References

Bakir, V., McStay, A., 2017. Fake news and the economy of emotions. Digital J. 1–22. https://doi.org/10.1080/21670811.2017.1345645.

Bar-Ilan, J., 2007. Google bombing from a time perspective. J. Comput.-Mediat. Commun. 12 (3), 910–938. https://doi.org/10.1111/j.1083-6101.2007.00356.x.

Bawden, D., Robinson, L., 2009. The dark side of information: overload, anxiety and other paradoxes and pathologies. J. Inf. Sci. 35 (2), 180–191. https://doi.org/10.1177/0165551508095781.

Bozdag, E., 2013. Bias in algorithmic filtering and personalization. Ethics Inf. Technol. 15 (3), 209–227. https://doi.org/10.1007/s10676-013-9321-6.

Brennen, B., 2017. Making sense of lies, deceptive propaganda, and fake news. J. Media Ethics 32 (3), 179–181. https://doi.org/10.1080/23736992.2017.1331023.

Brin, S., Page, L., 2012. Reprint of: the anatomy of a large-scale hypertextual web search engine. Comput. Netw. 56 (18), 3825–3833. https://doi.org/10.1016/j.comnet.2012.10.007.

Cozza, V., Hoang, V. T., Petrocchi, M., & Spognardi, A., 2016. Experimental Measures of News Personalization in Google News, Cham.

Datta, A., Datta, A., Jana, S., Tschantz, M.C., 2015. Poster: information flow experiments to study news personalization. Paper presented at the 2015 IEEE 28th Computer Security Foundations Symposium (CSF). IEEE.

Diakopoulos, N., 2015. Algorithmic accountability. Digital J. 3 (3), 398–415. https://doi.org/10.1080/21670811.2014.976411.

Dutton, W. H., Reisdorf, B. C., Dubois, E., & Blank, G., 2017. Search and Politics: The Uses and Impacts of Search in Britain, France, Germany, Italy, Poland, Spain, and the United States. Quello Center Working Paper No. 5-1-17. Retrieved from https://ssrn.com/abstract=2960697.

Epstein, R., Robertson, R.E., 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. Proceedings of the National Academy of Sciences of the United States of America 112 (33), 4512–4521. https://doi.org/10.1073/pnas.1419828112.

Finkle, T.A., 2012. Corporate entrepreneurship and innovation in silicon valley: the case of Google Inc. Entrepreneurship Theory Pract. 36 (4), 863–884.

Gillespie, T., 2014. The relevance of algorithms. In: Gillespie, T., Boczkowski, P., Foot, K. (Eds.), Media Technologies. MIT PRess, Cambridge, MA, pp. 167–194.

Gillespie, T., 2017. Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. Inf. Commun. Soc. 20 (1), 63–80. https://doi.org/10.1080/1369118X.2016.1199721.

Gomes, B., 2017. Our latest quality improvements for Search. Retrieved from https://www.blog.google/products/search/our-latest-quality-improvements-search/.

Google, 2016. Algorithms. Retrieved from https://www.google.com/insidesearch/howsearchworks/algorithms.html.

Haim, M., Graefe, A., Brosius, H.-B., 2018. Burst of the filter bubble? Digital J. 6 (3), 330–343. https://doi.org/10.1080/21670811.2017.1338145.

Hannak, A., Sapieyznski, P., Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C., 2013. Measuring personalization of web search. Paper presented at the Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, Brazil.

Harrison, G., 2015. Google, big data, and hadoop. In: Next Generation Databases: NoSQL, NewSQL, and Big Data. Apress, Berkeley, CA, pp. 21–37.

Harrison, X., 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ 2, e616. https://doi.org/10.7717/

peerj.616.

Hoang, V.T., Spognardi, A., Tiezzi, F., Petrocchi, M., De Nicola, R., 2015. Domain-specific queries and Web search personalization: some investigations. arXiv Preprint arXiv:1508.03902.

Höchstötter, N., Lewandowski, D., 2009. What users see – structures in search engine results pages. Inf. Sci. 179 (12), 1796–1812. https://doi.org/10.1016/j.ins.2009.01.028.

Horling, B., 2009. Personalized Search for everyone. Retrieved from https://googleblog.blogspot.be/2009/12/personalized-search-for-everyone.html.

Jarrett, K., Hillis, K., Petit, M., 2012. Google and the Culture of Search. Routledge, New York.

Just, N., Latzer, M., 2016. Governance by algorithms: reality construction by algorithmic selection on the internet. Media, culture and society 39 (2), 238–258. https://doi.org/10.1177/0163443716643157.

Netmarketshare, 2018. Search Engine Market Share. Retrieved from https://netmarketshare.com/search-engine-market-share.aspx.

Kammerer, Y., Gerjets, P., 2014. The role of search result position and source trustworthiness in the selection of web search results when using a list or a grid interface. Int. J. Human-Comput. Interact. 30 (3), 177–191. https://doi.org/10.1080/10447318.2013.846790.

Kang, H., McAllister, M.P., 2011. Selling you and your clicks: examining the audience commodification of Google. tripleC 9 (2), 141–153.

Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., & Mislove, A., 2015. Location, Location, Location: The Impact of Geolocation on Web Search Personalization. Paper presented at the Proceedings of the 2015 Internet Measurement Conference, Tokyo, Japan.

Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C., 2012. Explaining the user experience of recommender systems. User Model. User-Adap. Inter. 22 (4), 441–504. https://doi.org/10.1007/s11257-011-9118-4.

Langville, A.M., Meyer, C.D., 2006. Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton.

Lee, M., 2011. Google ads and the blindspot debate. Media Cult. Soc. 33 (3), 433–447. https://doi.org/10.1177/0163443710394902.

Moz, 2016. Google Algorithm Change History. Retrieved from https://moz.com/google-algorithm-change.

Napoli, P.M., 2014. Automated media: an institutional theory perspective on algorithmic media production and consumption. Commun. Theory 24 (3), 340–360. https://doi.org/10.1111/comt.12039.

Ørmen, J., 2016. Googling the news. Digital J. 4 (1), 107–124. https://doi.org/10.1080/21670811.2015.1093272.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., Granka, L., 2007. In Google we trust: users' decisions on rank, position, and relevance. J. Comput.-Mediat. Commun. 12 (3), 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x.

Pariser, E., 2011. The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think. Penguin.

Peacock, S.E., 2014. How web tracking changes user agency in the age of Big Data: the used user. Big Data Soc. 1 (2), 1–11. https://doi.org/10.1177/2053951714564228.

Penna, L., Quaresma, M., 2015. How We Perceive Search Engines. In: Marcus, A. (Ed.), Design, User Experience, and Usability: Interactive Experience Design: 4th International Conference, DUXU 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part III. Springer International Publishing, Cham, pp. 74–81.

Seymour, T., Frantsvog, D., Kumar, S., 2011. History of search engines. Int. J. Manage. Inf. Syst. 15 (4), 47–58.

Silverwood-Cope, S., 2012. Wikipedia: Page one of Google UK for 99% of searches. Retrieved from https://www.pi-datametrics.com/wikipedia-page-one-of-Google-uk-for-99-of-searches/.

Smyth, B., Coyle, M., Briggs, P., 2011. Communities, collaboration, and recommender systems in personalized web search. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (Eds.), Recommender Systems Handbook. Springer US, Boston, MA, pp. 579–614.

Steiber, A., Alänge, S., 2013. The formation and growth of Google: a firm-level triple helix perspective. Social Sci. Inf. 52 (4), 575–604. https://doi.org/10.1177/0539018413497833.

Vaidhyanathan, S., 2012. The Googlization of Everything: (And Why we Should Worry). University of California Press, Berkeley.

Van Couvering, E., 2007. Is relevance relevant? Market, science, and war: discourses of search engine quality. J. Comput.-Mediat. Commun. 12 (3), 866–887. https://doi.org/10.1111/j.1083-6101.2007.00354.x.

Zuiderveen Borgesius, F., Trilling, D., Moeller, J., Bodó, B., de Vreese, C.H., Helberger, N., 2016. Should we worry about filter bubbles? Internet Policy Review. J. Internet Regul. 5 (1). https://doi.org/10.14763/2016.1.401.