Digital Technologies and Selective Exposure: How Choice and Filter Bubbles Shape

News Media Exposure

Exposure to news media is increasingly mediated by digital technologies. Today, online media is the preferred news source in most advanced democracies, above TV and well above printed media. Moreover, two thirds (65%) of online media users prefer to use a side-door of some kind (e.g., social media, search engines, news aggregators) to access news outlets (Newman et al. 2017). These new ways of accessing news have shifted content curation from editorial boards to individuals, their social networks and algorithmic information sorting (Bakshy et al. 2015; Flaxman et al. 2016). Although many have seen in these technologies the potential for increasing exposure to diverse news media and opinions (e.g., Benkler 2006), the success of movements that appear to be immune to any factual evidence has reinvigorated claims concerning the potential of these technologies for personalizing information (Bruns 2017:1), locking people in echo-chambers (Sunstein 2009) and filter-bubbles (Pariser 2011).

Much recent research has contributed to this debate asking about the prevalence of selective exposure in audiences' online news consumption behaviour (Dubois and Blank 2018; Dvir-Gvirsman et al. 2016; Flaxman et al. 2016; Gentzkow and Shapiro 2010; Nelson and Webster 2017; Weeks et al. 2016). While these are valuable contributions, with very rare exceptions (e.g., Flaxman et al. 2016), they have not addressed the contribution of specific technologies (e.g., social media, search engines) on observed selective exposure nor the mechanisms underlying selectivity. Therefore, even if from current research we can conclude that there is very little evidence of selective exposure in online news consumption, we still do not know whether this is the outcome primarily of "choice" (or voluntary exposure) or of "filter-bubbles" (or involuntary exposure)[1].

In this paper, we inquire about these different mechanisms by looking at the role of different origins (direct visits, social media, and search engines) in selective exposure. We use a unique dataset that combines survey and online web-tracking data from Spain to study the impact on selective exposure of three different points of access to news: direct (1), referred from Facebook (2), and referred from Google (3). We focus on these platforms because they are the largest within their kind (i.e., social media and search engines) and they account for an important part of news consumption (see discussion below). Using a multi-level approach, we also analyse the impact of some individual attributes (e.g., ideology, political interest) whose influence in selective exposure is still unclear and explore cross-level interactions between these individual-level attributes and origins to news media.

We find that direct navigation increases selective exposure compared to most referred-based navigation —though the effect is small; that Facebook has no direct effect on selective exposure, and that Google considerably reduces selective exposure. More interestingly, we find that the relationship between points of access to news (e.g., direct, Facebook, Google) and selective exposure is strongly moderated by ideology, suggesting that search engines and social media are not content neutral. Our findings suggest a rather complex picture regarding selective exposure online.

Digital technologies and selective exposure

Selective exposure is a core concept in communication and media studies, which states that given a chance, individuals will chose to consume media that reinforces their previous beliefs (Klapper 1960; Katz and Lazarsfeld 1955). This behavior would be caused by cognitive dissonance (Festinger 1957) and would be responsible for the minimal effects of the media on changes in attitude and opinion (e.g., Lazarsfeld, et al. 1944).

According to mainstream accounts, technology may exacerbate selective exposure in the current media environment mainly through two mechanisms: choice and filter bubbles (Dubois and Blank 2018:731). Assuming that individuals have a preference for likeminded information, choice prompts selective exposure by giving individuals the opportunity to choose, among some range of diverse opinions, the one that best matches their previous beliefs (Iyengar and Hahn 2009; Mutz and Martin 2001). Hence, when individuals are given some choice, selective exposure is expected to be the outcome of voluntary action.

Filter bubbles in contrast refer to systems of information selection that are automatized and beyond individual control, notably, algorithmic curation (Bakshy et al. 2015; Dubois and Blank 2018; Zuiderveen Borgesius et al. 2016). Although algorithms are based on users' past choices and tastes, filter bubbles are seen to be conducive to selective exposure from involuntary action and without users' consent (Zuiderveen Borgesius et al. 2016, p.3). Some have depicted these two mechanisms in different ways. For example, Zuiderveen Borgesius et al. (2016) refer to choice as "self-selected personalization" and to filter bubbles as "pre-selected personalization". Here we use these concepts interchangeably.

Direct access to news. Under direct access to news media selective exposure is likely to be the outcome of choice or self-selected personalization. Direct visits to news media online should not be very different from old ways of accessing news through traditional media, in which newsgathering has typically been seen as the outcome of conscious choice (Tewksbury et al. 2001:533), and to rely on brand (or source) as the main heuristic criterion for information selection (Messing and Westwood 2012; Sundar et al. 2007).

There is considerable evidence of partisan selectivity when people are asked to choose among media labels that they associate with different (partisan) biases (e.g., Iyengar and Hahn 2009; Messing and Weswood 2012; Stroud 2008; Turner 2007). Based on these findings, many have claimed that the expansion of choice that is characteristic of the new

media environment will only exacerbate partisan selective exposure (e.g., Bennett and Iyengar 2008; Iyengar and Hahn 2009). According to these accounts, a high-choice media environment coupled with the proliferation of partisan outlets will lead partisan motivations to play a more active role in online news selection (Iyengar and Hahn 2009; Mutz and Martin 2001; Stroud 2008).

This view, however, has been contested by recent work. Data from real-life news consumption behaviour show that when consuming news outlets online people tend to rely on their favourite outlets, which tend be mainstream and centrist (Flaxman et al. 2016; Gentzkow and Shapiro 2010; Guess 2016). More importantly for our purposes, this behaviour — i.e., consuming mainstream, centrist outlets — has been shown to be the prevailing one when people access news directly (Flaxman et al. 2016).

In spite of these findings, there are several reasons to expect selective exposure under direct navigation. First, even if recent studies do not find evidence of selectivity in online news consumption, experimental and survey studies consistently find evidence of a confirmation bias in information selection (e.g., Garrett 2009b, 2009a; Hart et al. 2009; Iyengar and Hahn 2009; Knobloch-Westerwick and Meng 2009; Stroud 2008). Second, most of these studies do not discriminate between different technologies; hence, most of the diversity they find could be the result of accidental exposure to information, which might be most clearly promoted by search engines and social media (we discuss this below). Finally, most of the evidence is US-based and might be influenced by this country's media system. Some have argued that selective exposure is not only influenced by choice but also by the media system (Mutz and Martin 2001; Goldman and Mutz 2011). According to this view, a polarized pluralistic media system, such as the Spanish one (Hallin and Mancini 2011), would make it easier for people to exercise (partisan) selective exposure than a liberal system, such as the US.

Social media. Two mechanisms are at work in social media, which have the potential to personalize news in an involuntary way: homophily, which refers to the tendency that people have to connect with similar others (McPherson et al. 2001), and algorithms, which pre-select and recommend information to people based on their past choices and tastes. Although homophily involves choice (people choose their friends and peers in social media) it leads to unintended forms of exposure because users "transfer" news exposure decisions to friends who pre-select news stories for them (Bright 2016; Singer 2014).

Although most studies on echo-chambers and filters bubbles in social media have focused on Twitter (e.g., Barberá et al. 2015; Colleoni et al. 2014; Conover et al. 2011), here we focus on Facebook because it is by far the largest platform and the most frequently used for getting news. In 2017, 47% of Spanish web users declared using Facebook for getting news in the last week and a similar percentage of US users (47%) reported the same (Newman et al. 2017)[2]. Most studies focus in Twitter because data is easier to access. So an additional reason to study Facebook is that in spite of being the largest platform for getting news we know much less about the dynamics of news exposure in this platform.

Despite the potential for news personalization in Facebook, existing evidence does not support the filter bubble hypothesis (e.g., Bakshy et al. 2015; Zuiderveen Borgesius et al. 2016; Nechushtai and Lewis, 2019; Fletcher and Nielsen, 2018). In contrast, several factors may account for cross-cutting exposure in social media. First, even though social media reproduce homophily (e.g., Colleoni et al. 2014), they are also more likely to prompt weak ties (e.g., Barberá 2014), which increase opportunities to encounter novel information (Granovetter 1977; Prior 2008) and have been found to be fertile ground fostering cross-cutting exposure (Mutz and Mondak 2006). Second, even if social media are increasingly popular for getting news, given their socially dominant character, in these media people are more likely to encounter news while doing other things (e.g., Kim et al. 2013; Valeriani and

Vaccari 2015; Gottfried and Shearer 2016), and unintended exposure has been found to increase cross-cutting political exchanges and deliberation (Brundidge 2010; Wojcieszak and Mutz 2009). Finally, social media make available alternative cues for information selection — e.g. social recommendations— that may potentially override partisan cues (Knobloch-Westerwick et al. 2005; Messing and Westwood 2012; Nelson and Webster 2017). Moreover, using experimental design, Messing and Westwood (2012) have shown that given a choice people prefer relying on social recommendations rather than on partisan cues. They contend that information utility, which increases willingness to engage with attitude-discrepant messages (Knobloch-Westerwick and Kleinman 2012; Valentino et al. 2009), might explain this result. All these factors are more likely to be at work in platforms with a strong social character such as Facebook than in more specialized, news oriented platforms such as Twitter.

Search engines. The main mechanism with the potential for pre-selected personalization in search engines is algorithmic curation. Google and other search engines rely on algorithms for selecting among thousands of billions of information stored in the Internet, which they then present and recommend to the viewer. Although there are a number of search engines in the market (e.g., Yahoo, Bing, AOL…) to help users find information, Google is by far the search engine with the greatest market share worldwide. In Spain, Google's market share is 90.67%, and in the US its share is 78,81% (http://gs.statcounter.com). In addition, search engines (i.e., Google) have become increasingly popular platforms for accessing news. For 25% of the respondents in a study search engines are the preferred way of accessing news, and together search engines and social media are preferred gateways to news for almost 50% of web users (Newman et al. 2017). Recent research shows that Google search results may influence the evaluation of sources in many fields of behaviour, including voting (e.g., Epstein and Robertson 2015). So

worries about personalization are warranted not only because many rely on Google to be informed, but also because Google search results have been shown to directly affect political behaviour.

Despite these worries, research has found no evidence of substantial personalization in Google searches (e.g., Hannak et al. 2013; Haim et al. 2018; Hoang et al. 2015; Puschmann 2017; Fletcher & Nielsen, 2018; Nechushtai and Lewis, 2019). In contrast, several factors might help to explain cross-cutting exposure in Google. First, most people see search engines as a fair and unbiased source of information (Fallows 2005; Dutton et al. 2017), as neutral and ideologically blind (Sundar and Nass 2001) and neutral sources are associated with higher source credibility (Sundar and Nass 2001), which in turn might increase the likelihood of exposure to cross-ideological information by making users less defensive about media content (Druckman et al. 2012; Knobloch-Westerwick et al. 2015). Second, studies examining the use of heuristics in Google searches have found that rank order is by far the most important cue in information selection (Haas and Unkel 2017; Murphy et al. 2006; Chitika-Insights 2013) and heuristics that are not overtly partisan may increase opportunities for cross-ideological exposure (Messing and Westwood 2012). Third, Google might help users find sources that are not mainstream and contribute to diversify people's news media diets (Athey et al. 2017; Puschmann 2017).

From this discussion we would expect direct navigation to increase selective exposure compared to referred-based navigation and Facebook and Google to decrease it compared to direct navigation. However, we would also expect some interaction effects of news origins with individuals' political orientations, notably political interest and ideology. Politically-motivated individuals tend to have stronger opinions and more coherent views of the political world (Converse 1964; Zaller 1992), which might lead them to rely more often on partisan cues in their news consumption decisions (Barberá

and Sood 2014b; Iyengar and Hahn 2009). If, as some studies have found (e.g., Dutton et al., 2017), politically interested individuals are more likely to rely on popular news-referred platforms such as Google and Facebook to search political information, this might increase selective exposure in these platforms compared to direct forms of navigation.

Moreover, recent studies begin to debate whether news referred platforms such as Facebook and Google are content neutral (e.g. Hancock et al. 2018; Haim et al. 2018, Puschmann 2017; Dutton et al. 2017). Some US-based studies have found that liberals are more active in social media than conservatives (Batshy et al. 2015; Anderson and Jiang 2018). Moreover, a recent YouGov study found that in UK users perceive left-wing content to me more widespread in Facebook (Dahlgreen 2016). Following these works, we might expect ideology also to interact with origins in shaping selective exposure. In particular, based on existing works about Facebook (e.g. Dahlgreen 2016) we would expect Facebook and having a left-wing ideology to increase selective exposure compared to other forms of navigation.

To sum up, the following would be our testing hypotheses:

H1: Direct navigation will increase selective exposure compared to referred-based navigation.

H2: Exposure to news media through Facebook will decrease selective exposure compared to direct navigation.

H2.1: Facebook together with political interest will increase selective exposure compared to direct navigation

H2.2: Facebook together with being a leftist will increase selective exposure compared to other forms of navigation.

H3: Exposure to news media through Google will decrease selective exposure compared to direct navigation

> *H3.1: Google together with political interest will increase selective exposure compared to direct navigation.*

Data and measures

Design. This analysis is based on a combination of survey and online web tracking data for 408 individuals coming from a pool of the Spanish online population. Survey data was used to tap participants' left-right self-placement, political interest and other relevant socio-demographic traits. Web-tracking online data was used to trace participants' online navigation behaviour during a period of three-months, extending from January 26 to April 27, 2015, which has been estimated to give an accurate description of peoples' online media exposure habits (Athey et al. 2017). Navigation data enables us to rely on direct measures of exposure and also, given its granularity, to analyse the role of different origins to news media on selective exposure. For our study, we targeted individuals that consumed a minimum of information online[3] and who had their residence in all-Spanish regions except for Catalonia[4].

Panel. Participants in our study are part of an opt-in panel of a Spanish market-research firm, which worked with us on all aspects of the sample and the implementation of the survey. Recruitment was done using online contacts and offering incentives for completing structured questionnaires on their personal electronic devices (home computers, tablets, or cell phones). We targeted a sample of 1000 people and retained 40,8% of the subjects for our tracked sample (N=408). Hence, a total of 408 explicitly agreed to share their anonymized browsing history for our study, a figure that accords with previous studies analysing audience online news consumption (Guess, 2016; 2015).

Participants in our panel match the characteristics of the Spanish online population in most socio-demographics and relevant attitudes (Robles et al. 2012). They tend to be younger, more educated, and more politically interested than the average Spanish citizen, although ideologically they are slightly to the right of the average Spanish citizen (see the Online Appendix[5] for more details). To further check the representativeness of our sample, we compared the list of the top 20 most-visited news sites by our tracked sample with that provided by Alexa, one of the most widely used meters of online news audiences (Alexa Internet 2014), for the Spanish online population. We obtained a strong Pearson rank-order correlation score (.81). Hence, participants in our study tend to visit most of the same outlets (at least 80%), and with a similar frequency, as the general Spanish online population, testifying further to the representativeness of our sample with regards to a central element of our study such as online news consumption habits. In the Online Appendix we include additional analyses testing to the sample's representativeness.

Despite these similarities, we cannot make overgeneralizations from our final sample to the Spanish online population because people who voluntarily accept being tracked are generally less concerned about privacy. Yet we see this attitude as an advantage and assume that they will not modify their news consumption routines as a result of our study. Notably, our subjects agreed to being tracked long before we started the study, which may have also helped to mitigate any initial change in their regular behaviour.

Tracking data and coding procedure. We used a dataset on navigation data on desktop devices for a period of 92 days, which contains a total of 1,024,026 URLs of visited sites corresponding to 79,071 domains for 408 unique users. These observed data for online navigation was filtered in order to identify news exposure for the top 42 most-visited news outlets in Spain according to Alexa rankings[6]. Our final dataset is composed of the visits to the 42 top most-visited news outlets (our units of observation), which amount to 40,683

visits, and 3.97% of total site visits during our period of study. More information concerning the collection of the tracking data and its characteristics is provided in Online Appendix.

Measurements.

Selective exposure. The dependent variable in this study is selective exposure, which is defined at the visit level. To operationalize this variable we first had to classify outlets and individuals according to their ideologies (see below). Visits to news media that are consistent with the visitor's ideology are coded as 1, and as 0, otherwise. Table 1 includes the description of this and all other key variables.

Media Slant. To measure media slant we relied on reported media positions from the survey. Respondents provided the ideological position of each media outlet they usually visit (left, right, or neutral). We classified an outlet as partisan if at least 50% of respondents perceived it as right- or left-leaning. Our classification is shown in the Online Appendix (Table A5). In our classification, almost half of the news media included in our list of 42 are classified as partisan (M=0.51). More importantly, the list of partisan media includes the two major Spanish news outlets, *El Mundo* and *El Pais*, a result that matches previous classifications and research (Hallin and Mancini 2004, Newman et al., 2017). This is consistent with the characteristics of the Spanish media system, which has been classified as polarized pluralism (Hallin and Mancini 2004), in contrast, for example, to liberal media systems such as the US, where, according to previous research (Flaxman et al. 2016; Gentzkow and Shapiro 2011; Guess 2016), mainstream outlets and news portals tend to be classified as neutral, and partisan media are generally small in size.

Political leanings. To measure people's political leanings, we used ideological self-placement in a left-right scale from 0 (extreme left) to 10 (extreme right). Participants placing themselves in positions 6-10 of the scale are coded as right, and, those placing themselves in

positions 0-4, as left. Following other studies on selective exposure (e.g., Gentzkow and Shapiro 2011), moderates —those self-placed at 5 in the scale in our study— were eliminated from the analysis.

Origins. We consider a visit to be *direct* if there is no overlapping between the current and the previous session or if when visiting a new site the previous site session has been closed. More technically, for a visit to be coded as direct the difference between the starting time of a visit to outlet j *(SNt $_{i+1}$)* and the starting time of the previous visit *(SNt $_i$)* must be larger than the duration of navigation in the site of origin (we add an extra second to account for the opening and closing session of the site) or:

$$(SNt_{i+1}) - (SNt_i) > (DNt_i) + 1,$$

Conversely, we consider a visit to be *referred* if the opposite condition holds; that is, if the difference between the starting time of a visit to outlet j *(SNt $_{i+1}$)* and the starting time of the previous visit *(SNt $_i$)* is smaller than the duration of navigation in the site of origin. Hence, if the following condition holds we code a visit as referred[6]:

$$(SNt_{i+1}) - (SNt_i) < (DNt_i) + 1,$$

We consider a visit to a news outlet j to have its origin in Facebook if: (1) the site that immediately precedes the visit to j is Facebook and (2) it fulfils the condition of being referred[7]. The same coding procedure is applied for Google. All other indirect referrals (e.g. Twitter) have been coded as "others".

Political interest. Political interest is assessed by asking: "How much you would say you are interested in politics: Very much (3), quite interested (2), hardly interested (1), or not at all (0)?" The mean political interest in our panel is 1.97 (SD=0.76).

Education. We have included a third variable measured at the individual level that accounts for educational level. Given the skewed distribution of this variable (as the sample

was highly educated) we have coded it as having attended college against all the other possibilities.

Table 1 about here

The nature of our dependent variable requires a logistic estimation to test the relationship between origins and selective exposure. Yet, given the nested structure of our data (where visits are nested in individuals) we cannot use a simple logistic regression because we would be violating some fundamental assumptions of regression, as for example that observations must be independent and errors must not be correlated.

Using a multi-level approach allows us not only to explore the effect of individual (second) level variables (i.e., political interest, ideology) in the outcome variable (i.e., selective exposure), but also to explore cross-level interactions between first and second level variables, and thus, to model varying-intercept and varying-slope effects.

Results

To test our hypotheses we ran a series of hierarchical models. The results are displayed in Table 2. Table 2's first column reproduces a null model, followed by a model that only includes first level variables and a model that includes both first and second-level variables. The first model includes information for the variance of individual-level (Level 2 in our analysis) intercepts. Most importantly, the intra-class correlation (ICC) figure suggests that the amount of variation to explain at the visits level due to factors related to the upper level of analysis (individuals in our case) is as high as 34%. This is more than enough to justify a multilevel analysis.

The coefficients in model two read as regular logistic coefficients, estimated with proper standard errors. We ran the same models twice because our hypotheses are established against different categories of reference: referred-based (H1) and direct navigation (all the others). Here, we only show the results for the models using direct navigation as the reference category because it is the relevant reference category for most of the hypotheses (H2, H2.1, H2.2, H3, H2.1). However, the results with other referrals as the reference category can be found in the Online Appendix.

Models 2 and 3 show results for first-level and first and second-level predictors, respectively, assuming heterogeneous intercepts (different average levels of selective exposure across individuals). As we can see from these models, of all the origins, only Google has a significant effect on selective exposure and in the expected direction —i.e., it reduces the probability of selective exposure (H3). Neither Facebook nor other referrals have a significant effect on selective exposure compared to direct navigation.

Table 2 about here

Models 2 and 3, however, assume that the effects of origins on selective exposure are homogenous across individuals (i.e., they assume invariable slope effects), which may yield biased estimates if within slope effects are in fact not homogenous. When freeing the slope of the variable origin (Model 4) we see some changes, implying that the homogeneity (slope) assumption is probably wrong. First, the coefficient for Google increases, suggesting a larger Google effect. Second, we see for the first time a significant negative coefficient for other referrals. This suggests that other referrals have a negative effect on selective exposure, or symmetrically, that accessing news media directly increases the probability of selective exposure relative to arriving from

referrals other than Google and Facebook. This can be seen more clearly from Table A6 in the Online Appendix, which presents the same models with other referrals as the category of reference. From Model 3 in Table A6, we have estimated that direct navigation increases the probability of selective exposure by almost 3% at the maximum, compared to referred-based navigation. Hence, when we assume heterogeneous slope effects, we find a positive effect of direct navigation on selective exposure —albeit not a strong one— and some support for H1.

The fifth and sixth models show the cross-level interactions. As we can see from the coefficients and standard errors in Model 6, political interest is not a relevant moderator in the relationship between origins and selective exposure, lending no support to hypotheses H2.1 and H3.1. However, as Models 2-6 show, it has a strong direct effect in selective exposure, increasing the probability of selectivity by 14% at most. In contrast, left-right self-placement (and more particularly being a leftist) not only is a strong direct predictor of selective exposure but also comes out as a powerful moderator between most origins (e.g., Facebook and Google) and selective exposure. In particular, as expected (H2.2), among left-wing individuals, Facebook referred-news navigation increases the probability of selective exposure compared to direct navigation by 7%.

Figure 1 represents graphically the interaction included in Model 5 (Table 2). It plots predicted probabilities of selective exposure for different origins by ideological group.

Figure 1 about here

As we can see, for left leaners the predicted probability of selective exposure increases from 27% to 34% when the origin of news changes from direct access to

Facebook, while it drops to 17% when it changes to Google. We can also see that the opposite is true for right leaning individuals: the probability of selective exposure drops from 43% to 32% when the origin changes from direct access to Facebook, while it increases to 48% when it changes to Google. In other words, origins have opposite effects on left and right-wing ideologues: Facebook increases selective exposure among left leaners and decreases it among right leaners, while Google decreases it among left leaners and increases it among right leaners.

Discussion

Numerous works have studied the prevalence of selective exposure in online news consumption. Yet, they have rarely addressed the contribution of news origins to observed selective exposure. This has hindered our ability to learn about the consequences of different pathways to news for selective exposure, and about the mechanisms underlying online selectivity. Relying on a unique dataset that combines survey and web-tracking online data from Spain, this paper has analysed the impact of different origins to news media and of several moderators in an attempt to uncover a complex picture about selective exposure online and to bring light on the mechanisms underlying this phenomenon.

With regards to the (direct) impact of origins, we have expected direct navigation to increase selective exposure compared to referred-based navigation (H1) and Facebook and Google to reduce selective exposure compared to direct navigation (H2 and H3). Yet, we have also expected political orientations, such as interest in politics and ideology, to interact between origins and selective exposure. In particular, we have expected political interest to increase selective exposure in popular news-referred platforms such as Facebook and Google (H2.1 and H3.1) compared to direct navigation, and having a

leftist ideology to increase selective exposure in Facebook compared to other forms of

accessing news (H2.2).

Our study partially supports our expectations. Although we did not find support for

H1 when we assumed homogenous slope effects, when we allowed the effects of origins

(the slope) to vary across individuals, we found that direct navigation increases selective

exposure, even though the effect is small (at most it increases the probability of selective

exposure by approximately 3%). As for the role of specific platforms, we found that

Facebook has no direct effect in selective exposure but Google is a significant actor

shaping online selective exposure. We estimated that Google reduces selective exposure at

most 9%, which is a considerable effect. In line with other studies (e.g., Fletcher and Nielsen,

2018) we show that Google not only does not lock people into filter bubbles but it increases

media consumption across ideological lines.

However, our most interesting results probably concern the cross-level

interactions. Although we did not find political interest to moderate origins and

selective exposure, we found a strong interaction effect between origins and ideology. In

particular, we found ideology and news-referred platforms (i.e., Facebook and Google)

to shape selective exposure in opposite ways —Facebook increased the probability of

selective exposure among left-leaners and decreased it among right-leaners and Google

decreased the probability of selective exposure among left-wing ideologues and

increased it among right-wing ideologues. These results suggest at least two

explanations.

A first explanation relates to group behaviour: left and right leaners might behave

differently as active information seekers in these platforms. For example, conservatives might

use Google to search directly the newspaper they normally read while liberals might rely on

search terms more often. In social media, left leaners might be more active news-consumers

than right leaners, prompting greater selectivity as consumption intensifies (Prior 2013; Guess 2016).

The second explanation relates to information availability in these platforms and suggests that neither Facebook nor Google might be neutral in information retrieving (e.g., Haim et al. 2018; Hancock et al. 2018; Puschmann 2017). There is some evidence that Facebook might be biased to the left in content retrieving because left-wing individuals are over represented in this platform (Mellon and Prosser 2017) and they are also more active in it (Anderson and Jiang 2018). Also, some works are beginning to discuss and study the neutrality of Google in content retrieving, suggesting that it might have an ideological bias (e.g., Haim et al. 2018; Hancock et al. 2018; Puschmann 2017). Although evidence is scarce and it is (still) too soon to know, the strong interaction effect that we find between ideology and news-referred platforms suggests that these platforms might not be content neutral, something that given the increasing popularity of these platforms as news levers and their potential effects on voting behaviour (Epstein and Robertson 2015) and should deserve much more attention from research in the future.

Finally, we found that political interest and ideology are strong direct predictors of selective exposure. In line with other studies (e.g., Barberá et al. 2015), we found that ideology matters to account for patterns of news media exposure and that left-leaning news consumers are significantly more likely to consume media from the other side of the political spectrum than right-leaning individuals. In particular, we estimated a probability change of at most 20% due to ideological leaning. In line with previous studies (e.g., Stroud 2011), we also found that political interest increases —not decreases— the probability of selective exposure. This suggests that politically interested people not only are more capable of

identifying the content of political messages ex ante but also they are more likely to rely on this information (i.e., partisan cues) to select political information.

This study suffers from several limitations. First, exposure is measured at the media not at the content level. Second, exposure is observed only through desktops, leaving out news access through mobiles. Although access to news through mobiles has experienced the most important increase in recent years, desktops continue to be the leading device in online news consumption (Newman et al. 2018). Finally, measurement of news media exposure is limited to the top 42 most-visited outlets according to Alexa, which might leave out consumption of niche media. In our favour, it is worth noting that in line with other studies (e.g., Gentzkow and Shapiro 2011) we found that among our panellists those that consumed niche media were also consumers of mainstream media.

In spite of these limitations, this study addresses the contingencies of online selective exposure through the use of unobtrusive data and a multi-level approach. It shows that more nuanced approaches are needed to tackle the complexity of selective exposure and opens a promising line of analysis for uncovering some of the contingencies influencing selectivity online.

Endnotes

[1]In what follows, we use the term choice to refer to voluntary exposure and the term filter bubbles to refer to involuntary exposure resulting mainly (but not only) from algorithmic curation (Zuiderveen et al., 2016; Dubois & Blank, 2018).

[2]Users reporting to get news from Facebook dropped in the US from 47% to 39% in the last year (Newman et al., 2018).

[3] To guarantee a maximum of useful information, we aimed at selecting people for our panel who consumed a minimum of online information. We used four filter questions–from

less to more restrictive. The less restrictive question asked people if they followed current events or read news online. A positive answer in any of the four questions was enough to include the individual in our panel. Only four people were filtered out, which proves the inclusiveness of our questions and rules out potential problems of selection bias.. The questionnaire can be accessed here: https://www.dropbox.com/sh/61j9own413ho1oq/AACzPaFJ5lql8yLJrPLBA0gsa?dl=0.

[4] We excluded Catalonia from our study to keep the analysis as simple and tractable as possible. In contrast to Spain, where all issues collapse to a single left-right dimension and politics can be characterized as "one-dimensional", in Catalonia (as well as in the Basque country) politics are multidimensional, in that more than one dimension structures political conflict. Adding dimensions would have unnecessarily complicated a study about segregation. Size and politics informed our exclusion of Catalonia. Catalonia represents 20 % of the Spanish population. In addition, during the period of study, the second dimension was especially salient in Catalonia, as testified by the rise of the Catalan pro-independence movement, which increased its support from 14% in 2006 to 41% in 2015 (Muñoz and Tormos 2015).

[5] The Online Appendix is available here: https://www.dropbox.com/s/s9zlbg768sib6vf/Online_Appendix.pdf?dl=0.

[6] We selected the top 42 most-visited news outlets reported by Alexa because they represent an important percentage of the total audience for news according to both Alexa and ComScore. The strong correlation (0.906) between media's position in both these lists—Alexa and ComScore—proves the accuracy of our media sample. Media outlets that have not been included in this list are niche news providers that yield relatively low audience-reach figures according to ComScore. Our list of 42 media includes 99,85% of all reported visits to online news outlets in our sample, which adds to its exhaustiveness.

⁷ Note that session overlapping is only a necessary condition for referral, not a sufficient one—e.g., the user might have the previous session open without this necessarily implying that he/she has arrived to the next site through this (previous) one. Thus, even though session overlapping is the best measure at hand for referral, it is an imperfect one because we cannot be certain that a site—even if there is session overlapping—has actually worked as a referral for the next site.

References

Alexa Internet, I. 2014. About the Alexa Traffic Rankings. Retrieved July 14 from www.alexa.com/help/traffic-learn-more.

Anderson, M., and Jiang, J. 2018. "Liberal Democrats more likely than other groups to be politically active on social media". *Factank*, November 5. http://www.pewresearch.org/fact-tank/2018/11/05/liberal-democrats-more-likely-than-other-groups-to-be-politically-active-on-social-media/. (November 2, 2018).

Athey, S., Mobius, M. M., and Pál, J. 2017. "The impact of aggregators on internet news consumption". Working Paper, Stanford University.

Bakshy, E., Messing, S., and Adamic, L. A. 2015. "Exposure to ideologically diverse news and opinion on Facebook". Science 348(6239): 1130–1132.

Barberá, P. 2014. "How social media reduces mass political polarization. Evidence from Germany, Spain, and the US". Job Market Paper, New York University, 46.

Barberá, P., and Sood, G. 2014. Follow Your Ideology: A Measure of Ideological Location of Media Sources. Unpublished manuscript.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. 2015. "Tweeting from left to right: Is online political communication more than an echo chamber?" Psychological science, 26(10):1531–1542.

Benkler, Y. 2006. The wealth of networks: How social production transforms markets and freedom. Yale: Yale University Press.

Bennett, W. L., and Iyengar, S. 2008. "A new era of minimal effects? The changing foundations of political communication". Journal of Communication 58(4):707–731.

Bright, J. 2016. The Social News Gap: "How News Reading and News Sharing Diverge". Journal of Communication 66(3):343–365.

Bright, J. 2018. Explaining the emergence of political fragmentation on social media: the role of ideology and extremism. Journal of Computer-Mediated Communication 23(1):17–33

Brundidge, J. 2010. "Encountering "difference" in the contemporary public sphere: The contribution of the Internet to the heterogeneity of political discussion networks". Journal of Communication 60(4):680–700.

Bruns, A. 2017. "Echo chamber? What echo chamber? Reviewing the evidence". In 6th Biennial Future of Journalism Conference (FOJ17), 14-15 September 2017, Cardiff, UK. (Unpublished).

Chitika-Insights 2013. The Value of Google Result Positioning.

Colleoni, E., Rozza, A., and Arvidsson, A. 2014. "Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data". Journal of Communication 64(2):317–332.

Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., and Flammini, A. 2011. "Political Polarization on Twitter". ICWSM 133:89–96.

Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. In Ideology and Discontent, ed. David Apter. New York: Free Press.

Dahlgreen, W. 2016. Facebook seen as a left-wing echo chamber. *YouGov*, May 12. https://yougov.co.uk/topics/politics/articles-reports/2016/05/12/facebook-left-wing-echo-chamber. (May 7, 2017).

Druckman, J. N., Fein, J., and Leeper, T. J. 2012. "A source of bias in public opinion stability". American Political Science Review 106(02):430-454.

Dubois, E., and Blank, G. 2018. "The echo chamber is overstated: the moderating effect of political interest and diverse media". Information, Communication and Society 21(5):729–745.

Dutton, W., Reisdorf, B., 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

Dvir-Gvirsman, S., Tsfati, Y., and Menchen-Trevino, E. 2016. "The extent and nature of ideological selective exposure online: Combining survey responses with actual web log data from the 2013 Israeli Elections". New Media and Society 18(5):857–877.

Epstein, R., and Robertson, R. E. 2015. "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections". In Proceedings of the National Academy of Sciences 112(33):4512–4521.

Fallows, B. 2005. Search Engines Users. Pew Research Center. http://www.pewinternet.org/2005/01/23/search-engine-users.

Festinger, L. 1962. A theory of cognitive dissonance. Stanford: Stanford university press.

Flaxman, S., Goel, S., and Rao, J. 2016. "Filter bubbles, echo chambers, and online news consumption". Public Opinion Quarterly 80(S1):298-320.

Fletcher, R., & Nielsen, R. K. (2018). "Automated serendipity: The effect of using search engines on news repertoire balance and diversity". Digital Journalism, 6(8), 976-989.

Freedman, J. L., and Sears, D. O. 1966. "Selective exposure". Experimental Social Psychology 2:57.

Garrett, R. K. 2009a. "Echo chambers online?: Politically motivated selective exposure among Internet news users". Journal of Computer-Mediated Communication 14(2):265–285.

Garrett, R. K. 2009b. "Politically motivated reinforcement seeking: Reframing the selective exposure debate". Journal of Communication 59(4):676–699.

Gentzkow, M., and Shapiro, J. M. 2011. "Ideological Segregation Online and Offline". The Quarterly Journal of Economics 126(4):1799-1839.

Granovetter, M. S. 1977. "The strength of weak ties". Social networks 2:347–367.

Haas, A., and Unkel, J. 2017. "Ranking versus reputation: perception and effects of search result credibility". Behaviour and Information Technology 36(12):1285–1298.

Haim, M., Graefe, A., and Brosius, H.-B. 2018. "Burst of the filter bubble? Effects of personalization on the diversity of Google News". Digital Journalism 6(3):30–343.

Hancock, J., Metaxa-Kakavoul, D., and Park, J. 2018. "Are Google search results politically biased?". *The Guardian*, Setember 6. https://www.theguardian.com/commentisfree/2018/sep/06/google-search-results-rigged-news-donald-trump. (Febrary 25, 2019).

Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., and Wilson, C. 2013. "Measuring personalization of web search". In Proceedings of the 22nd international conference on World Wide Web 27–538.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., and Merrill, L. 2009. "Feeling validated versus being correct: a meta-analysis of selective exposure to information". Psychological bulletin 135(4):555.

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

Iyengar, S., and Hahn, K. S. 2009. "Red media, blue media: Evidence of ideological selectivity in media use". Journal of Communication 59(1):19–39.

Katz, E., and Paul, F., and Lazarsfeld, J. 1955. Personal Influence. The Part Played by People in the Flow of Mass Communication. New York: Routledge.

Kim, Y., Chen, H.-T., and De Zúñiga, H. G. 2013. "Stumbling upon news on the Internet: Effects of incidental news exsure and relative entertainment use on political engagement". Computers in Human Behavior 29(6):2607–2614.

Knobloch-Westerwick, S., and Meng, J. 2009. "Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information". Communication Research 36(3):426-448.

Knobloch-Westerwick, S., Johnson, B. K., and Westerwick, A. 2015. "Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts". Journal of Computer-Mediated Communication 20(2):171-187.

Knobloch-Westerwick, S., and Kleinman, S. B. 2012. "Preelection selective exposure: Confirmation bias versus informational utility". Communication Research 39(2):170–193.

Knobloch-Westerwick, S., Sharma, N., Hansen, D. L., and Alter, S. 2005. "Impact of popularity indications on readers' selective exposure to online news". Journal of broadcasting and electronic media 49(3):296–313.

Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). The people's choice: How the voter makes up his mind in a presidential election. New York and London: Columbia University Press.

McPherson, M., Smith-Lovin, L., and Cook, J. M. 2001. "Birds of a feather: Homophily in social networks". Annual review of sociology 27(1):415–444.

Messing, S., and Westwood, S. J. 2012. "Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online". Communication Research 41(8):1042-1063.

Mellon, J., and Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, *4*(3), 2053168017720008.

Murphy, J., Hofacker, C., and Mizerski, R. 2006. "Primacy and recency effects on clicking behavior". Journal of Computer-Mediated Communication 11(2):522–535.

Mutz, D. C., and Martin, P. 2001. "Facilitating communication across lines of political difference: The role of mass media". American Political Science Review 95(1): 97-114.

Nelson, J. L., and Webster, J. G. 2017. "The Myth of Partisan Selective Exposure: A Portrait of the Online Political News Audience". Social Media+ Society 3(3):22-46.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A., and Nielsen, R. K. 2017. Reuters Institute Digital News Report 2017.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A., and Nielsen, R. K. 2017. Reuters Institute Digital News Report 2018.

Nechushtai, E., and S. C. Lewis. (2019). "What Kind of News Gatekeepers Do We Want Machines to Be? Filter Bubbles, Fragmentation, and the Normative Dimensions of Algorithmic Recommendations". Computers in Human Behavior 90: 298-307.

Pariser, E. 2011. The filter bubble: What the Internet is hiding from you. London: Penguin.

Prior, M. 2008. "Are hyperlinks "weak ties"". In The hyperlinked society: questioning connections in the Digital Age, 227–249.

Singer, J. B. 2014. "User-generated visibility: Secondary gatekeeping in a shared media space". New Media and Society 16(1):55–73.

Stroud, N. J. 2008. "Media use and political predispositions: Revisiting the concept of selective exposure". Political Behavior 30(3):341–366.

Stroud, N. J. 2011. Niche news: The politics of news choice. Oxford: Oxford University.

Sundar, S. S., Knobloch-Westerwick, S., and Hastall, M. R. 2007. "News cues: Information scent and cognitive heuristics". Journal of the American Society for Information Science and Technology 58(3):366–378.

Sundar, S. S., and Nass, C. 2001."Conceptualizing sources in online news". Journal of Communication 51(1):52–72.

Sunstein, C. R. 2009. Republic. com 2.0. Priceton: Princeton University Press.

Turner, J. 2007. "The messenger overwhelming the message: Ideological cues and perceptions of bias in television news". Political Behavior 29(4):441–464.

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., and Tucker, J. A. 2016. "Of Echo Chambers and Contrarian Clubs: Exposure to Political Disagreement Among German and Italian Users of Twitter". Social Media+ Society, 2(3): 98-114.

Valentino, N. A., Banks, A. J., Hutchings, V. L., and Davis, A. K. 2009. "Selective exposure in the Internet age: The interaction between anxiety and information utility". Political Psychology 30(4):591–613.

Valeriani, A., and Vaccari, C. 2015. "Accidental exposure to politics on social media as online participation equalizer in Germany, Italy, and the United Kingdom." New Media and Society 18(9):1857-1874.

Weeks, B. E., Ksiazek, T. B., and Holbert, R. L. 2016. "Partisan enclaves or shared media experiences? A network approach to understanding citizens' political news environments". Journal of Broadcasting and Electronic Media 60(2):248–268.

Wojcieszak, M. E., and Mutz, D. C. 2009. "Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement?" Journal of communication 59(1):40–56.

Zaller, J. 1992. The Nature and Origins of Public Opinion. Los Angeles: Cambridge University Press.

Zuiderveen Borgesius, F., Trilling, D., Moeller, J., Bodó, B., de Vreese, C. H., and Helberger, N. 2016. "Should we worry about filter bubbles?" Internet Policy Review 5(1): 56-98.

Table 1. Summary of variables

	mean	sd	min	max
Selective exposure	0.37	0.48	0	1
Origin: Direct	0.59	0.49	0	1
Origin: Facebook	0.05	0.22	0	1
Origin: Google	0.07	0.26	0	1
Origin: Other	0.28	0.45	0	1
Left	0,74	0.43	0	1
Political Interest	1.98	0.77	0	3
College	0.65	0.48	0	1

Table 2: Multilevel logistic estimation of selective exposure.

			Dependent variable:			
			Selective exposure			
	(1)	(2)	(3)	(4)	(5)	(6)
Origin: Facebook		0.087	0.088	0.102	-0.449**	-0.139
		(0.064)	(0.064)	(0.100)	(0.185)	(0.281)
Origin: Google		-0.284***	-0.286***	-0.351***	0.216	-0.352
		(0.056)	(0.056)	(0.097)	(0.160)	(0.263)
Origin: Other(referrals)		-0.035	-0.034	-0.108*	-0.182	-0.173
		(0.033)	(0.033)	(0.063)	(0.119)	(0.171)
Ideology(left)			-0.724***	-0.719***	-0.690***	-0.719***
			(0.170)	(0.182)	(0.180)	(0.182)
Political interest			0.561***	0.562***	0.563***	0.538***
			(0.103)	(0.103)	(0.102)	(0.109)
College			-0.074	-0.059	-0.062	-0.073
			(0.155)	(0.156)	(0.155)	(0.156)
Facebook*Left					0.764***	
					(0.215)	
Google*Left					-0.830***	
					(0.195)	
Other(referrals)*Left					0.100	
					(0.139)	
Facebook*Pol.Int						0.127
						(0.134)
Google*Pol.Int						-0.003
						(0.127)
Other(referrals)*Pol.Int						0.033
						(0.082)
Constant	-0.882***	-0.851***	-1.327***	-1.346***	-1.367***	-1.291***
	(0.084)	(0.085)	(0.252)	(0.257)	(0.256)	(0.266)
N(grp)	300	300	300	300	300	300
ICC	.34	.34	.32			
Intercept Std.Dev	1.29	1.29	1.23	1.24	1.24	1.24
Random slope	No	No	No	Yes	Yes	Yes
Observations	32,013	32,013	32,013	32,013	32,013	32,013
Log Likelihood	-16,172.730	-16,158.110	-16,136.730	-15,887.160	-15,871.060	-15,886.680
Akaike Inf. Crit.	32,349.460	32,326.220	32,289.470	31,808.320	31,784.110	31,813.360
Bayesian Inf. Crit.	32,366.210	32,368.080	32,356.460	31,950.680	31,959.960	31,980.830

Note: *p<0.1; **p<0.05; ***p<0.01

Figure 1: Interaction between origin and ideology