Media Exposure and Opinion Formation in an Age of Information Overload

Simon Munzert, Humboldt University of Berlin
Pablo Barberá, University of Southern California
Andrew Guess, New York University
JungHwan Yang, University of Wisconsin-Madison

Abstract

We present the outline of a major online panel tracking study that aims to investigate media exposure and opinion formation in an age of information overload. The overarching project asks if online sources and social media platforms exacerbate polarization, and inequality in political knowledge and behavior, and what role weak ties play in moderating citizens' information diets. To that end, we study how offline events are translated into media coverage, why people decide to consume or avoid coverage, and how these decisions affect attitudes and behavior. We combine passive metering technology to capture the online media consumption of a representative sample of individuals in Germany and the United States, draw on machine learning and natural language processing methods to estimate the topic and ideological slant of each separate piece of content consumed, and in parallel directly survey the panelists at regular intervals to monitor changes in issue attention, opinion, and political knowledge.

1 Introduction

Compromise is the foundation of a healthy democracy. Many worry, however, that the ingredients necessary to foster it are no longer features of the modern-day media ecosystem: Social media flattens individuals into avatars of a worldview. Expressions of solidarity and outrage go viral, while more nuanced arguments remain in obscurity. It is easier to find information that supports rather than challenges one's ideological predispositions. Such normative concerns are common in both the popular press and in scholarly research about the effects of internet connectivity on political polarization. And while it is clear that the ubiquity of new communication technologies is profoundly changing the way we experience the social world, the pessimism of the emerging consensus tends to rely on evidence subject to a dizzying number of selection and measurement difficulties. This proposed research project addresses these difficulties by harnessing the very technological advances that have so profoundly altered our information environment in the 21st century.

Coming to terms with the vast effects wrought by a continuous stream of new and potentially surprising information requires grappling with a host of challenging methodological problems. First, who is exposed to what information? This measurement problem is one of the most vexing challenges in social science, made even more urgent by our increasingly multifaceted and fragmented online information environment. Second, what drives this media consumption? Most attempts to answer this question are plagued by reverse causation. And third, how do different types of people behave in this context? Despite an explosion of large-scale social media data, there is a tendency to use aggregate data to make inferences about individuals.

This project uses an interdisciplinary approach to solve each of these shortcomings in the existing research. To measure exposure to information, we will use tracking software to passively capture the online media consumption of a representative sample of individuals in the United States and Germany. We leverage the exogenous nature of certain types of events—e.g., mass shootings (in the United States) and unexpected political developments (during the 2017 election campaign in Germany)—in order to identify the direction of causality. Finally, to understand how events can have different effects on people, we build a large panel survey in both countries linking individuals' responses to events, their demographic characteristics, and media consumption (via mobile, web, and social media).

We match this design with computational technologies for estimating the ideological slant of each separate piece of content—an article or blog post, for example—consumed

by our panelists. To do so, we will take advantage of recent developments in natural language processing and machine learning to project massive quantities of text data onto an informative, low-dimensional scale. This final component provides the link between the people, the events they pay attention to (or ignore), and the ways in which these events are mediated through social networks and online news sources. It is the intersection of all three that will allow us to shed new light on whether the internet is fundamentally bringing us together as societies, or driving us apart.

This manuscript reviews the proposed design and method of a three-year project on news consumption and political communication in a fragmented media age. In the next section, we outline our theoretical framework. We then summarize a number of proposed "sub-projects" related to the overarching theme. Next, we provide details on our research design and our methodological approach before concluding.

2 Theoretical Framework

The popularization of the internet as a communication tool represents a dramatic shift in citizens' media diet. According to data from the Pew Research Center, the internet has become the second most important source of information for U.S. adults, and a majority (62%) of them see news on social networking sites such as Twitter or Facebook on a daily basis (Gottfried and Shearer, 2016). The goal of this project is to understand how this transformation affects (1) media coverage of political issues and events, (2) the process driving individuals' consumption of political information, and (3) how this consumption in turn affects political attitudes, preferences, knowledge, and behavior. In our project, we will examine how two competing views of the ways in which the internet transforms exposure to news—broader media availability and choice, and social consumption of political information—affects each of these three phases of individuals' media consumption process.

Most existing scholarly accounts of how the internet transforms citizens' media diet have focused on the first view. The common underlying argument is that the increased availability and diversity of information to which citizens have access online broadens media choice, which should lead to audience fragmentation and selective exposure (Sunstein, 2001; Prior, 2007). In this context, individuals who are not interested in politics can avoid receiving any political information, since it is argued that the likelihood of incidental exposure to any political content decreases (Prior, 2005). On the contrary, citizens with

higher levels of interest in politics, who are more likely to have strong partisan identities, can choose to increase their consumption of political news, and to do so in a way that helps them avoid information that may challenge their viewpoints (Levendusky, 2013).

This common wisdom in previous studies is at odds with how most citizens consume political information on the internet, which increasingly takes place through social networking sites. Our complementary argument focuses on the fact that these platforms generate "social consumption" of political information (Kaplan and Haenlein, 2010). Unlike news portals, blogs, or online forums, the political information to which citizens are exposed through platforms like Facebook or Twitter is what their friends, family, co-workers, and acquaintances decide to produce or share. Even if news organizations and journalists are also present on social media sites, most users report receiving political messages from other individuals in their personal networks.

The social consumption of news represents a consequential change in how citizens keep up with current political events. Since individuals are now inadvertently exposed to the news their friends and acquaintances decide to share (Brundidge, 2010), selective exposure to ideologically congenial information decreases (Lazarsfeld, Berelson and Gaudet, 1944; Sears and Freedman, 1967). As Messing and Westwood (2012) demonstrate, friends' recommendations are powerful social cues that reduce the role of partisanship in news consumption. Their findings illustrate how individuals are likely to click through and read news stories shared by their friends and acquaintances even if they potentially disagree with the message of that story. Most importantly, since individuals do not self-select the content they are exposed to, even those with low levels of political interest will be potentially exposed to political news, in a clear contrast with the general intuition that the internet allows citizens to "tune out" from political information. This transformation is likely to have important implications for political behavior. On one hand, more exposure to political news can increase levels of political knowledge and interest. However, social networking sites facilitate communication with weak ties (Bakshy, Messing and Adamic, 2015), which tend to be more ideologically heterogeneous, and thus may induce "cross-pressures" that may depress political participation (Mutz, 2002*a*,*b*).

Each of these competing frameworks generates different predictions regarding the effects of the internet as a source of news on the three components of news processing cycle: coverage, attention/exposure, and attitudes/behavior (see Figure 1). These effects can take place at two interconnected levels: individual and societal. We are interested in studying how this transformation affects political knowledge, ideological extremism,

3

Figure 1: Overview of theoretical framework and research questions
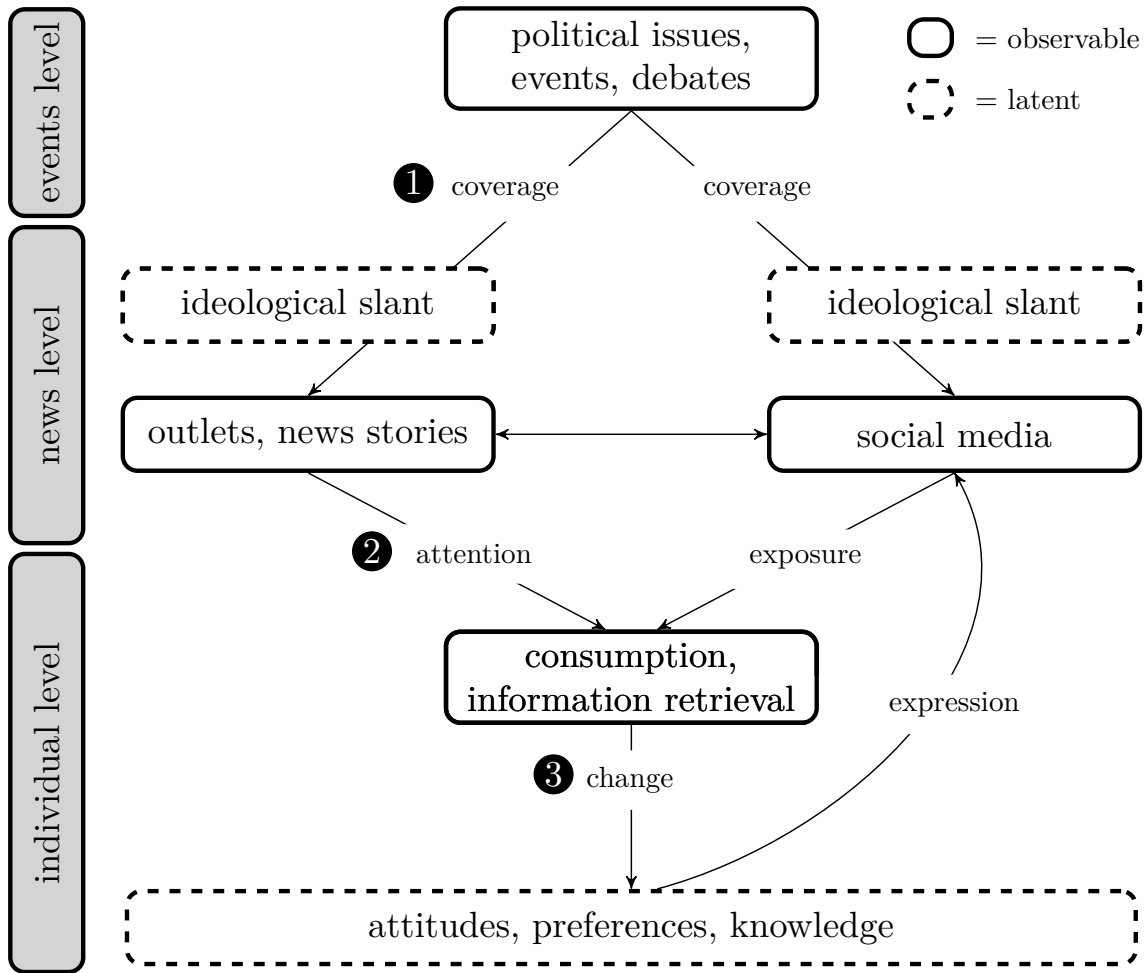
and opinions on issues; but also macro-level agenda-setting dynamics and mass political polarization. The aim of this project is to provide empirical evidence to adjudicate between predictions generated for each view at these two levels of aggregation.

3   Subprojects

3.1   How the media covers politics

The first step of the analysis corresponds to how events and political issues are translated into content that is published in media outlets or shared on social media. Depending

on which of our two contending accounts is predominant, we would expect to observe different outcomes regarding *what* is covered (more emphasis on issues vs. events), *how* it is covered (more emphasis on policy vs. politics), and *who* receives more coverage and is capable of influencing the public agenda (top-down vs. bottom-up information flow).

Based on our main claim that social consumption of information is now the driving force explaining exposure, we expect to observe the following three empirical patterns:

Hypothesis 1a	News outlets pay more attention to events as opposed to issues.
Hypothesis 1b	News outlets focus more on politics as opposed to policy.
Hypothesis 1c	Ordinary citizens have a greater degree of influence on the issues covered by news outlets.

The main factor explaining these hypotheses is related to how the consolidation of the internet as a news source affects the expected audience of political information. In a context of broader media availability, we would expect news consumers to be a self-selected sample of the population that exhibits greater interest in partisan politics and higher levels of political sophistication. However, as social networking sites become the predominant method for news consumption, the audience of political information is increasingly becoming more representative of the general population. In this scenario, news outlets face incentives to produce content that appeals to the mass public and that can be more prone to social sharing, given that a large portion of content consumed through social media originates from traditional media outlets (as represented by the arrow between these two sources in Figure 1). As a result, we expect them to emphasize content that is more likely to be socially shared and "go viral", adding a new dimension to the literature on news values (Lippman, 1922; Galtung and Ruge, 1965; Staab, 1990; Donsbach, 2004). This type of content, often derided as "clickbait", will be predominantly composed of breaking news events and stories about partisan conflict, and presented in such a way that can induce physiological arousal to increase its social appeal (Berger, 2011; Berger and Milkman, 2012).

We expect the magnitude of these shifts to depend on the extent to which media markets are audience-driven and the degree of social media penetration. In countries where most media outlets are privately owned and when they face strong pressures to produce "viral" content, our assumption is that news outlets as suppliers of information want to satisfy the demand of consumers and favor entertainment over substantive but less appealing information or arguments. This pressure is not as strong in other countries

where media outlets are publicly owned and thus are thought to serve an educating purpose, or where social media penetration is lower.

One of the defining features of social media websites is the abundance of user-generated content, which co-exists with content shared from traditional news outlets and competes for users' attention. In this context, previous studies have demonstrated the complexity of predicting the popularity of news stories, in practice generating an environment in which "everyone's an influencer" (Bakshy et al., 2011). This change is part of a broader trend toward citizen journalism (Goode, 2009; Lewis, Kaufhold and Lasorsa, 2010) in which traditional news outlets cede part of their gatekeeping and agenda-setting power in exchange for access to hyperlocal news and cheaper content. We claim that one empirical implication of this transformation is a greater degree of influence of ordinary citizens on the events and issues that are covered by news outlets and receive attention on social media. In other words, we expect the agenda-setting to increasingly follow a bottom-up process (McCombs and Shaw, 1972).

3.2 Attention and exposure to political information

The question of how the internet affects media consumption patterns has received considerable scholarly attention (Bennett and Iyengar, 2008; Chadwick, 2006; Prior, 2013*b*; Tewksbury, 2003; Mutz and Young, 2011; Webster, 2014). A central debate in this literature is whether the increased availability and diversity of online media sources is contributing to a trend toward more ideologically homogenous media diets, similar to the effect of the introduction of cable news (Prior, 2007). Empirical evidence of persistent ideological sorting in online communication networks (Adamic and Glance, 2005; Conover et al., 2012; Colleoni, Rozza and Arvidsson, 2014) has been taken to suggest that internet use may exacerbate this trend. As Sunstein (2001) or Hindman (2008) argue, the internet may create communities of like-minded individuals where cross-ideological interactions and exposure to political diversity are rare, facilitating individuals' self-selected exposure to congenial media content (Garrett, 2009; Prior, 2007; Stroud, 2008). However, a more recent body of work using online consumption data has found relatively limited evidence of this phenomenon, with many Americans exhibiting quite balanced media diets (Gentzkow and Shapiro, 2011; Flaxman, Goel and Rao, 2016; Guess, 2016). Our project aims to contribute to this burgeoning literature.

However, two limitations remain. The first one is methodological: developing accurate and comprehensive measures of media consumption and its ideological slant at the

individual level has proven elusive in an increasingly complex media environment. We address this limitation in the next section. The second one is substantive: previous studies did not take into account how the success of social media platforms is transforming online media consumption.

Individuals now devote an increasing amount of their time online to social media platforms such as Twitter or Facebook, whose features are very different from traditional news websites or online forums. What previous studies did not pay enough attention to is that these platforms generate social consumption of political information and create opportunities for individuals to be inadvertently exposed to political content. At the same time, these websites facilitate connections with weak ties, such as acquaintances, non-close friends, and distant relatives (Bakshy et al., 2012; Centola and Macy, 2007; Granovetter, 1973), who tend to hold more ideologically heterogeneous positions (see e.g. Mutz, 2006, p.27).[1] As a result, we have the following expectations:

Hypothesis 2a	Most citizens are exposed to a higher degree of heterogeneity in their consumption of political information online than offline.
Hypothesis 2b	Most citizens are exposed to a higher degree of heterogeneity in their consumption of political information through social media than through other internet sites.

3.3 Attitudes, preferences, and knowledge

Understanding the process by which individuals consume and process political news is relevant because of its implications on their political attitudes and knowledge. Here we focus on three outcome variables that are likely to be affected by this transformation in political communication patterns, both at the individual and societal levels: ideological extremism (and mass political polarization), opinions on issues (and public opinion stability), and political knowledge (and the quality of public discourse).

The question of whether mass political polarization is on the rise has received broad scholarly and journalistic attention in recent years (see e.g. Abramowitz and Saunders,

---

[1]Some scholars, such as Pariser (2011) and Tufekci (2016), argue that the use of algorithms to rank the content that will appear in social media users' news feeds constrains the likelihood of chance encounters with diverse views, since these algorithms only select content that users are more likely to "like"; Pariser (2011) defines this situation as "filter bubble." However, these theoretical accounts have not been validated by empirical studies, which find high levels of exposure to diverse views on social media (Bakshy, Messing and Adamic, 2015; Barberá, 2014).

2008 and Fiorina and Abrams, 2008). A growing body of work argues that the introduction of the internet as a relevant communication tool is contributing to this trend (Farrell, 2012), because it creates "echo-chamber" environments where individuals are only exposed to like-minded political views that reinforce their previous ideological positions (Jamieson and Cappella, 2008). It may also encourage affective distaste for the out-party (Lelkes, Sood and Iyengar, 2015). However, if the hypotheses we presented in the previous sections are valid, we would be more likely to observe the opposite empirical pattern, particularly in contexts where social media penetration is higher. As a result of more cross-cutting exposure to information, we expect individuals to develop "greater awareness of rationales for oppositional views" (Mutz, 2002*b*, p.114), which should lead to political moderation (Allport, 1954) and to weaker partisan identities. We would also expect to observe lower levels of affective polarization (Iyengar, Sood and Lelkes, 2012) due to the discovery that social ties in individuals' networks have different perspectives.

A more general version of this argument would apply not only to ideological positions but also to positions on specific issues. If it is true that individuals are only exposed to information that reinforces their beliefs, we should see little to no change in their opinions, since any conflicting information that generates cognitive dissonance will be ignored as a result of motivated reasoning processes (Taber and Lodge, 2006). However, in an environment in which individuals are overloaded with cross-cutting information that is embedded within social cues (Messing and Westwood, 2012), we may expect opinions to be more fluid, and more prone to changes as a result of Bayesian updating processes (Barabas, 2004; Bartels, 1993; Gerber and Green, 1999; Bullock, 2009; Guess and Coppock, 2017).

It is often believed that increased availability of political information due to the consolidation of the internet as a communication tool should lead to a more informed citizenry (see, e.g., Kenski and Stroud (2006)), although the empirical evidence up to this point is contradictory (Prior, 2005). One potential explanation for this pattern could be conceptual: the internet may be increasing levels of factual knowledge, since it allows citizens to develop their "transactive memory" (Wegner, 1987), that is, their ability to easily retrieve information; but not necessarily their *civic* knowledge, which would comprise information about how political institutions work, the importance of voting, or the cognitive skills and habits required to engage in meaningful political action. If it is indeed the case that news outlets pay more attention to events and politics rather than issues and policy, we would

8

expect this second trend to be predominant, particularly so in countries with higher levels of social media penetration and lower baseline levels of civic engagement.

Because of these theorized processes, we have developed the following hypotheses:

Hypothesis 3a   Exposure to diverse views in the online consumption of political information induces political moderation.

Hypothesis 3b   Exposure to diverse views in the online consumption of political information induces changes in opinions about relevant political issues.

Hypothesis 3c   Exposure to political information online decreases levels of measured political knowledge.

3.4   The role of headlines in driving attention and political knowledge

During the 2016 U.S. election campaign, it became standard for commentators (especially on Twitter) to criticize news organizations for declining to clearly label statements as false in headlines and in social media posts. The implicit theory is that even if the content of a news story makes clear that a politician is telling untruths, this information may not be conveyed effectively to segments of the audience that do not click through to a story. Given the rapidly changing nature of social news consumption (outlined above), it seems plausible that many people learn about political developments—at least in broad outline— by scanning headlines in their social media feeds (and, increasingly, on their mobile phones). How these emerging habits distort people's understanding is only beginning to be understood (Dunaway & Searles N.d.). We plan to study the relationship between headlines, news content, and political knowledge in a series of experiments embedded in our research design. This is necessary in part because of potentially severe selection effects: For example, younger people may be more likely to get news on the Twitter mobile app.

On the macro level, there are parallel questions about whether news outlets present the same stories in systematically different ways on different platforms given assumptions (or perhaps knowledge driven by internal analytics) about different news audiences.

The hypotheses below focus on the interaction of partisanship, knowledge, and selective exposure. News organizations have an incentive to maximize traffic and therefore want to avoid deterring those for whom a story's conclusion may be counter-attitudinal. However, this may come at the cost of failing to fully inform those who have lower levels of

political interest and are therefore less likely to click through to a piece of political content in the first place.

Hypothesis 4a	Explicitly labeling political statements as "false" or "lies" in headlines or social media posts will induce selective exposure to a story, causing partisan divergences in knowledge.
Hypothesis 4b	Refraining from such upfront judgments will reduce *partisan* selectivity in consumption but strengthen the relationship between political interest and selectivity.
Hypothesis 4c	The labeling strategy causes people for whom the judgment is pro-attitudinal to update their beliefs about the target of the story.
Hypothesis 4d	The non-labeling strategy causes people who click through to the story to update their beliefs regardless of political predispositions.
Hypothesis 4e	The labeling strategy will more effectively increase knowledge among those with lower political interest via incidental exposure.

4 Data and Research Design

Any research designed to identify mechanisms of attention-building and information-based attitude change has to consider that the flow of information from political issues and events to individuals' reactions is both a dynamic and highly selective process. Not all information that is newsworthy actually makes it into the news, and not all information that is reported receives the same level of attention. Further, people get deliberately or inadvertently exposed to information and opinion. Reconsidering Figure 1, selection mechanisms are at work at each of the stages: Looking at event characteristics, various aspects of news value might play role for editors and journalists, who decide whether an event or topic is covered or not (Galtung and Ruge, 1965). At the news level, the ideological leaning of journalists but also peers in social networks affects whether particular news items are shared and how they are commented. Finally, at the individual level, both consumption filters (different people consume different media) and susceptibility filters (not all people are affected in the same way by the same information) are likely to be at work.

Our approach to the challenge of identifying mechanisms behind selective attention allocation and opinion formation builds upon a combination of three data collection

Figure 2: Overview of data collection setup
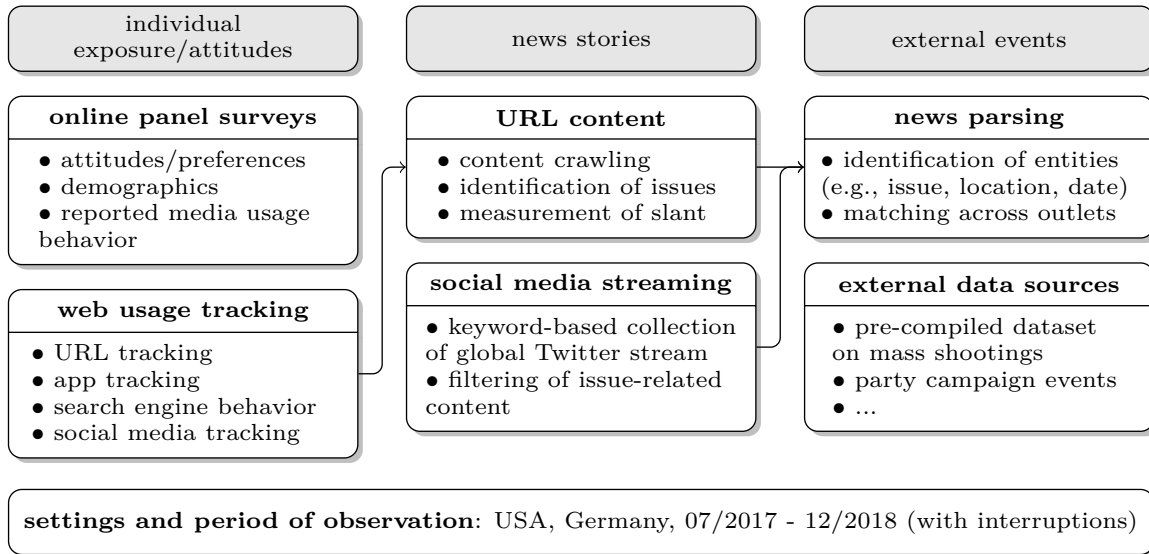
strategies, as illustrated in Figure 2. The first is the implementation of an online panel, which enables us to measure attitudes and political behavior over a longer time span for a representative sample of the population. In addition, by employing web tracking technologies on the panelists we will be able to observe individual exposure to news and information shared by social peers, which is often believed to be an important factor explaining political attitudes and behavior (Barberá et al., 2015; Berelson, 1954; Bond et al., 2012; Sinclair, 2012). The second strategy is to identify both the content and ideological slant of consumed information. Therefore, we will crawl content from online newspapers and blogs that has been consumed by the individuals observed in the panel. We will also gather content from social media, which helps us track what people are exposed to when using these platforms. The third strategy is to draw on existing and assemble new ground-truth datasets of relevant external events and conditions for a set of specific issues. Matching this data with information on media coverage and individual-level indicators of news consumption and opinion enables us to study factors driving (1) media coverage, (2) individuals' consumption, and (3) change in political attitudes and behavior. In the following, we will outline core elements of our research design in more detail.

4.1 Case selection: Germany and the United States

The project will consider two empirical settings, Germany and the United States. As is clear from the discussion of the state of the literature, the United States is the most thoroughly studied case with regards to our questions of interest. Our goal is to build on existing research and make our results comparable to previous findings. On the other hand, outside the U.S. there is remarkably little evidence to what extent partisan selective exposure exists and what role traditional and new media play in this regard (see, however, Bos, Kruikemeier and de Vreese, 2016, for a recently published study that investigates the role of public service broadcasting in the Netherlands). This includes the German case, which makes our project a valuable contribution by offering a comparative perspective on news attention and opinion dynamics. Most importantly, though, the two cases offer variation on a set of contextual factors we expect to play an important role for media consumption behavior and news exposure.

A core societal characteristic of the United States in general and American politics in particular is the vast gap between liberals and conservatives (or Democrats and Republicans) that is represented by the notorious red-blue map (DiMaggio, Evans and Bryson, 1996; Fiorina and Abrams, 2008), which also has left its marks in the mass public (Baldassarri and Gelman, 2008; Prior, 2013*b*). On the other hand, Germany has been described as a more consensus-oriented and comparatively depolarized society where people do not align along multiple issues (Munzert and Bauer, 2013). This distinction is relevant because we expect cross-cutting media exposure to be greater in countries with lower levels of mass political polarization.

Media markets in both countries also differ substantively. It has been theorized how characteristics of media systems affect mechanisms of media exposure and opinion formation (Hallin and Mancini, 2004). While public broadcasting continues to play an important role in Germany (in particular on TV but also online), the United States is market-dominated across news media. Therefore, we will be able to provide evidence on the extent to which it makes a difference if people are exposed to media that tends to prioritize infotainment news to generate revenue on the one hand, or rely on sources that attach more weight to the actual news value of information.

Another potentially important difference is the variation in social media penetration. According to data from the Pew Research Center, as of 2015 65% of US adults use social networking sites, and a large majority of them (63%) report getting news on these sites (Barthel et al., 2015). In contrast, in Germany as of 2014 only 50% of adults reported

using social networking sites, despite levels of internet penetration (79%) being comparable to those in the U.S., according to data from the Autumn 2014 Eurobarometer. This can affect to what extent people can willfully rely on (or are inadvertently exposed to) information spread by their social peers.

Apart from these persistent differences, there are also issue-related factors that make both cases appealing settings for our study. As for the German case, we expect the next federal election—which is commonly considered to be the most salient election—to take place in Fall 2017. Election campaigns are a particularly attractive case to study attention and opinion formation dynamics because candidates and party platforms constantly try to advertise their positions and, on the other hand, people seek out information informing their vote choice. That is, campaigns provide a setting where the fight over attention for various political issues is particularly intense. We plan to cover large parts of the campaign and a period in the aftermath of the election with our panel instrument (see below). Analogously, we plan to cover the campaign period and aftermath of the 2018 midterm elections in the U.S. This will give us leverage to compare attention-building dynamics in two highly relevant but distinct cases.

4.2 Surveying people's media consumption behavior and attitudes

4.2.1 Implementation of an online web tracking panel

Giving valid answers to the questions raised above requires fine-grained measures of media exposure at the individual level, as well as indicators of people's interest in political issues, attitudes and preferences toward policies. Further, studying cause-and-effect relationships of media exposure, attention building, and attitude change requires a longitudinal perspective. This might be the main reason why few, if any, comparable efforts exist to combine data of these kinds. It was not until very recently that predominantly online forms of media consumption—and related technologies that enable us to track online behavior—have made possible a unique opportunity to study the mechanisms of interest. We directly draw on these developments in our research design. The main component of our data collection and measurement efforts will be an online web tracking panel that provides three crucial features:

1. Continuous individual-level passive metering of web usage on desktop and mobile devices over a long time span;

Figure 3: Overview of panel design
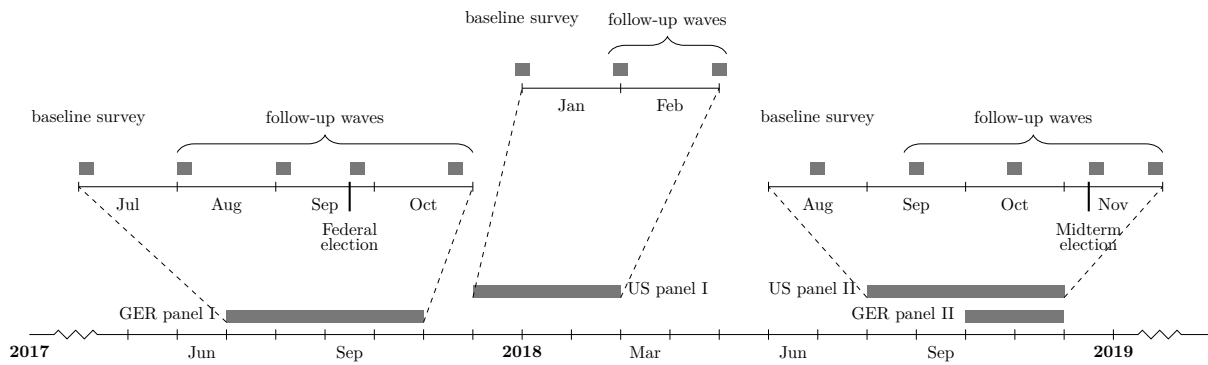
2. Repeated surveying of these respondents' political attitudes, issue priorities and other traits of interest;

3. Partially synchronized cross-country comparison.

Figure 3 illustrates the basic logic of the panel design. Both the German and the U.S. panels are designed as an interrupted panel that is split into two parts of differing lengths. The German panel will be launched roughly two months before the federal election date (September 2017) and cover the campaign period as well as the aftermath. The U.S. panel will launch in early 2018 to cover the primaries to the midterm elections. The second tracking period will start in August 2018 and cover the months through November 2018 when the midterm elections are scheduled in the U.S. As for the German panel, we will only cover the months of October and November during that year.

Both panels are made up of several waves. We will start with the first baseline survey in Germany in July 2017. In several follow-up waves, we will embed repeated measures of issue attention, knowledge about topical issues and events, and opinion on these issues. Given that we can draw on existing demographic data on the panelists from the provider, these follow-up surveys can be very short (5 minutes). In addition, we will save space in the questionnaires for questions dedicated to track ex-post evaluations of unexpected events. Within each period of observation, web usage on desktop and mobile devices, and potential exposure to social media feeds, is tracked continuously for the survey respondents. We will implement the passive metering component of the design via the YouGov Pulse panel, which enables tracking of people's web usage on desktop and mobile devices.

14

4.2.2 Sampling

Given ever-declining response rates on traditional telephone surveys, there is a growing urgency to identify reliable methods of achieving representative samples via online recruitment.[2] Surveys run by YouGov combine purposive sampling with a multi-stage sample-matching and weighting procedure (Rivers, 2006). The weights are generated using the joint distribution of known population characteristics—variables typically used by traditional polling companies such as demographics, but also party identification and political interest. The distribution is modeled by combining several high-quality data sources. These sources are combined to create a large synthetic sampling frame. A stratified sample is drawn from this frame and matched as closely as possible to YouGov's longstanding panel (with over a million members).

Sampling for the panel surveys in each country will follow a version of the procedure outlined above. Since our Pulse panels will be a subset of YouGov's traditional survey panels, we will undertake an additional round of sample matching to account for the extra level of selection. As part of our agreement with YouGov, the company will devote substantial efforts to recruit a fresh panel from scratch in both countries, ensuring that a matched sample representative of each population will comprise our study's respondents. For our panel, YouGov guarantees a net count of 1,000 respondents over the entire time span for each country. Accounting for usual panel attrition, the initial panels start with a higher gross number of respondents. Both we and YouGov expect to develop a set of best practices on methods for ensuring that hard-to-reach subgroups can be included (and retained) in Pulse panels.

4.2.3 Passive metering technology

Wakoopa, the tracking software used by YouGov, runs in the background of panelists' devices and collects anonymized visit data.[3] There are no technological limits to the types of websites that can be included in the data. Moreover, the software tracks web traffic[4] for all browsers installed on a user's computer. The technology does not slow the performance

---

[2]A recent benchmarking study by Pew found that among 9 opt-in web panels (including YouGov) and one probability panel, YouGov performed the best on a range of metrics (see http://www.pewresearch.org/2016/05/02/evaluating-online-nonprobability-surveys/). YouGov is "Sample I." Importantly, it performed the best on "civic benchmarks" including various forms of political participation, and it did the best in accurately characterizing subgroups.

[3]For a recent study employing a three-week sample of Pulse data in the United States, see Guess (2016).

[4]Passwords and financial transactions are ignored.

of users' computers and is transparent about the data that is being sent: Panelists can see a list of the last several captured URLs and can also pause tracking for 15 minutes. Of course, they can also uninstall the software at any time. Currently, YouGov encourages its panelists to install the software on as many devices as possible, including laptops, mobile phones, and tablets. The capabilities for mobile tracking are somewhat more limited for privacy reasons, but data on domain-level visits and app use will be collected.

4.2.4 Social media data

In the baseline survey at the beginning of the tracking period, we will ask respondents for their consent to match survey data with the content of their social media feeds. We want to be able to follow what people are exposed to on Twitter as well as the public content they produce themselves on the platform. Specifically, we will ask each respondent to share his or her Twitter handle if available. Additionally, we are developing an application that uses the Facebook API to collect data on respondents' likes and sharing behavior. This can also be installed with permission and provides data for up to three months after installation. Thus, in principle, we will be able to collect data on online media exposure and sharing behavior on the web, smartphones, and via both Twitter and Facebook accounts for the same subjects over an extended period of time while simultaneously gauging their attention to issues and political attitudes at multiple points. We turn to the survey design for this latter component next.

4.2.5 Assembling data on external events

As a third component of our data collection strategy, we want to make use of data on external events that are linked to political issues. This will give us methodological and substantial leverage. For instance, mass shootings regularly fuel the debate on gun control. In most cases, it can be safely assumed that shooting incidents are exogenous to the debate about these shootings and possible consequences. Therefore, they are a valuable piece of information that is not *per se* targeted at one ideological camp or a particular subgroup of the population. In doing so, we will be able to separate factual information on news events from news reports that are potentially affected by ideological slant or other editorial decisions. Legewie (2013) has suggested a way to exploit exogenous events—in his case, a terrorist attack taking place while a survey is in the field—as a quasi-experimental setup to examine its impact on related issue preferences. We plan to draw on this strategy and extend it to other issues and related events. Finally, a methodological benefit of collecting

data on events independent of news coverage is that it provides a way to match news sources (e.g., articles in different newspapers) according to the content they cover. People who consume different types of media still might become equally informed about specific issues or events, and data on external events provides a source to validate this claim.

Identifying a baseline set of external events is a major methodological challenge. Often all we know about events stems directly from news reports. However, using news reporting as the sole basis, as previously done in some applications (e.g., Atkinson, Deam and Uscinski, 2014) is not without problems, as this disregards the process of filtering and news selection done by journalists and editorial teams. Another popular strategy in linking external events to media frames and public attentiveness is to focus on rare and extreme events, such as ecological disasters (Birkland, 1997), war onset (Krosnick and Brannon, 1993), or terrorist attacks (Legewie, 2013). In principle, the study of the impact of such events seems attractive because, due to their nature as unexpected and sudden shocks, they frequently add a lot of new information to a debate and significantly shape the public as well as the political agenda. On the other hand, also almost by nature, such events tend to be very rare and idiosyncratic, which makes it hard to generalize on other settings (see Birkland, 1997). Furthermore, less can be learned about who pays attention if everybody pays attention.

For the above reasons, we employ a combination of strategies: In parallel to our web tracking and surveying efforts, we collect data on events that are closely linked to political issues and policies. For mass shootings, we can draw on own data collection efforts (see below) that can be extended for future periods as well as relying on multiple existing databases of event data (e.g., Stanford Mass Shootings of America project, 2016, and the FBI database). One particularly promising source is the Mass Shooting Tracker,[5] a crowdsourced database of shooting incidents, which does not rely solely upon editorial judgments that can lead to selection bias. For our analysis of attention building and attitude change during election campaigns, we will identify campaign-related events on party and candidate platforms. Therefore, we will crawl campaign websites and social media channels and classify published events by date, location, issues involved, news and social media coverage, among others. Finally, to capture major events which cannot be anticipated and, partly due to that reason, arouse interest among media and the public, we will parse documents and metadata from traditional and new media. Various

---

[5]See http://massshootingtracker.org/about

event detection techniques have been proposed to automate this process, e.g., by mining Twitter (Weng and Lee, 2011) or Wikipedia data (Osborne et al., 2012).

4.2.6 Survey design

To make our hypotheses on attention towards political information, attitudes, preferences, and political behavior testable, we need respondents' reports on some of these political variables. In particular, we will ask for turnout and vote choice at previous elections, political interest, political knowledge, partisanship, and ideological preferences.

As a core part of the survey, we will also monitor perceived attention, attitudes and preferences toward selected political issues. To understand how issues attract public attention and how people select information sources among a number of different channels that are available on the internet, social media, and other local, national, and international news media, our project turns to a variety of issues. We acknowledge that people's media use could differ by the nature of issues they are looking for. For example, people rely more on familiar and ideologically congruent sources when they look for political news (e.g., campaign news coverage) (Stroud, 2008; Prior, 2013b), whereas the preference for particular news sources might not be so strong when they look for event-driven news (e.g., natural disasters or terrorist attacks). Thus, our project considers a number of issues ranging from political to event-driven issues as well as other unexpected topics that might be pushed on the agenda during the field period. The selection of issues is motivated by a set of factors, including whether they are relevant in the political debate (such as election campaigns), polarized along societal or partisan lines (such as climate change policies), linked to external events (such as gun control), or provide local variation in salience (such as interest rates and gas prices).

In addition to two specific election campaigns, as detailed above, we will focus on mass shootings as quasi-random events that exogenously drive news coverage. Mass shootings unfortunately occur quite frequently in the U.S. Over 130 mass shootings that involved four or more victims happened during the last four years (January 2011-December 2015).[6] Gun control is often proposed (and contested) as a policy solution in the aftermath of such incidents, and they serve as useful focal points for observing how public attention to these issues evolves over time.

Our general approach will be to select a sample of issues to include in our baseline survey in both countries in the hopes that a subset of them will become salient in the

---

[6]Source: http://www.gunviolencearchive.org/

future. In this way, we will be able to track respondents' evolving attention dynamics and attitudes using both "pre-treatment" and "post-treatment" survey measures (and minimizing potential demand effects). Moreover, covering exogenous events — and retaining the ability to deploy last-minute "emergency waves" if necessary — will also provide us with maximum causal leverage in order to identify the mechanisms at work at each level of selection in the newsgathering/attention/consumption process.

5 Methodological Innovations

5.1 Measuring media exposure

Most previous studies of media effects rely on self-reported measures of media exposure. There are several main sources of error in this measurement approach. First and most clearly, imperfect recall is an important driver of people's general tendency to overreport their exposure to news (Prior, 2009). Also, social desirability may play a role since staying informed about political affairs might be seen as a signal of good citizenship. An additional set of issues more specific to online media is the sheer number of available sources and the variety of ways in which individuals can come into contact with media content, from bookmarks and email to social media and mobile apps (Guess, 2015). Due to these problems and others, Prior (2013*a*) recommends discarding survey-based approaches in favor of automatic monitoring of survey respondents' media use, as we propose here.

5.2 Measuring media slant

The other challenge of measuring media exposure is identifying the political bias of news sources. The lack of valid and reliable measures of news slant makes it hard to understand people's political information use and the effect of partisan news exposure. There are three basic approaches in previous literature: audience-based methods that use aggregate data on visitors to infer the political orientation of outlets (Golbeck and Hansen, 2014; Flaxman, Goel and Rao, 2016), content-based methods that use elements of media content—including the text of articles—to estimate the ideological position of publications (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010; Puglisi and Snyder, 2015; Budak, Goel and Rao, 2016), and methods using revealed preferences on social networks to scale both individuals and media outlets on the ideological spectrum (Barberá, 2015; Bond and Messing, 2015). Each method has its own advantages and disadvantages, but

most importantly for our purposes, one limitation currently shared by all three is the assumption that individual pieces of content (i.e., articles or blog posts) share the orientation of the media outlet. Operationally, this means that estimates are at the web domain level rather than at the web page level. This is a strong assumption because many media outlets carry a significant amount of ideologically heterogeneous content—for example, *The New York Times* features several conservative commentators in its opinion section, and CNN has adopted the confrontation between extreme views in the format of "panels" as one of the signature aspects of its news coverage.

Our emphasis in this project is to build and evaluate algorithms that identify the slant of political news at the URL level (instead of the domain level) using a combination of natural language processing, topic modeling, and traditional content analysis. We define slant as the position of the main (or average) views expressed in a news article on a latent multidimensional policy (or ideological) space.[7] We plan to test two different methods. One consists on supervised machine classification: After building a training dataset with URLs manually labeled by crowd workers, we will estimate different machine learning classifiers to predict the categories to which new articles belong to. These categories can be news or non-news content; left-wing, centrist or right-wing content; and even topic or political issue. The second method we plan to pursue builds upon our previous work estimating the political ideology of social media users (Barberá, 2015): Under the assumption that users will tend to share content that affirms their existing beliefs, we plan to test whether the average political ideology of Twitter users who share a specific URL is a good estimate of its political content.

We expect this innovation to contribute not only to the understanding of the main research questions in our project, but also the broader literature on media exposure and political communication.

5.3 Data processing and analysis

In combining data from multiple sources, covering two countries along several panel waves, the proposed project is exceptionally data-intensive in terms of volume, velocity, and variety.

---

[7]Note that there may still be some heterogeneity within articles (e.g. if they feature quotes by politicians from different parties), but we argue that the article is probably the minimal unit with which we can work using existing computational methods.

Based on ongoing work conducted by one of our team members (Guess, 2016), we anticipate that the URL-level dataset (second column in Figure 2) could contain as many as 50 million URL items per country. An important computational challenge in our project will be the development of an efficient pipeline that allows us to harvest the content of these URLs shortly after they have been consumed, process them to extract the relevant text data, and store it in a convenient format for analysis in cloud-based services that can facilitate collaboration while preserving the integrity and security of the dataset.

In the analysis stage of our project, we will rely on recent developments in large-scale statistical learning (Hastie et al., 2005) and distributed processing systems such as MapReduce (Dean and Ghemawat, 2008) to develop measures of the media slant of individual news articles. Using crowdsourcing techniques (Benoit et al., 2016), we will build training datasets that will then be used to implement machine learning classifiers that allow us to categorize articles as described above. To identify selection mechanisms and effects of (new) media exposure and exogenous information on opinions and preferences, we will combine these computationally demanding measurement techniques with recently suggested identification strategies from the causal inference toolbox (Barabas and Jerit, 2009; Legewie, 2013; Feezell, 2016).

We will also develop succinct summary measures to capture the diversity, density, and emotional content—in addition to the overall ideological lean—of subjects' Twitter feeds. We will continually scrape the feeds using the account handles provided to us by respondents.

6   Conclusion

From a normative standpoint, an informed society is believed to be an important component of a healthy democratic system (Delli Carpini and Keeter, 1996). From this perspective, understanding how individuals allocate their attention to existing media content, how they process this information, and how it impacts their political preferences and behavior is of utmost importance. This is particularly so given the many open questions about this process in the current literature about these topics. The increasingly complex media environment in which citizens are now embedded requires new and rigorous research to understand how profound transformations in political communication—such as the advent of the internet and social networking platforms as sources of political information—are transforming these processes (Bennett and Iyengar, 2008; Mutz and

Young, 2011). Crucially, developing informed policy proposals that aim at increasing political knowledge and civic engagement also require a better understanding of how individuals process political news in the internet age.

Our project introduces two major substantive innovations. First, we qualify the widely accepted theory of the internet as an "echo chamber" by arguing that inadvertent exposure to ideologically heterogeneous information through social media may actually increase cross-cutting interactions, with important implications for the study of political attitudes and behavior. Second, we provide a more nuanced analysis of these implications by examining processes both at the individual and at the aggregate level, and by studying how contextual variables may affect the strength of these changes.

The goal of our research is also to take a decisive step forward in the study of media consumption through the use of computational methods to provide more granular measures. In this sense, our methodological contribution is the development of machine learning methods that will allow us to estimate the ideological slant of the information that individuals consume and to apply these methods to a uniquely comprehensive dataset comprising this information, which we will be able to collect using web tracking technology. Our hope is to contribute to this literature by providing new data and methods that can be widely used across different disciplines to address longstanding questions in the study of political communication and behavior.

References

Abramowitz, Alan I. and Kyle L. Saunders. 2008. "Is polarization a myth?" *The Journal of Politics* 70(02):542–555.

Adamic, L.A. and N. Glance. 2005. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on link discovery*. ACM pp. 36–43.

Allport, Gordon Willard. 1954. *The nature of prejudice*. Addison-Wesley.

Atkinson, Matthew D., Maria Deam and Joseph E. Uscinski. 2014. "What's a Dog Story Worth?" *PS: Political Science and Politics* 47(4):819–823.

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow and Lada Adamic. 2012. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*. ACM pp. 519–528.

Bakshy, Eytan, Jake M. Hofman, Winter A. Mason and Duncan J. Watts. 2011. Everyone's an influencer: quantifying influence on Twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM pp. 65–74.

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

Baldassarri, Delia and Andrew Gelman. 2008. "Partisans without constraint: Political polarization and trends in American public opinion." *American Journal of Sociology* 114(2):408–446.

Barabas, Jason. 2004. "How deliberation affects policy opinions." *American Political Science Review* 98(04):687–701.

Barabas, Jason and Jennifer Jerit. 2009. "Estimating the Causal Effects of Media Coverage on Policy-Specific Knowledge." *American Journal of Political Science* 53(1):73–89.

Barberá, Pablo. 2014. "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S." *Unpublished Manuscript* .

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1):76–91.

Barberá, Pablo, Ning Wang, Richard Bonneau, John T Jost, Jonathan Nagler, Joshua Tucker and Sandra González-Bailón. 2015. "The Critical Periphery in the Growth of Social Protests." *PloS ONE* 10(11):e0143611.

Bartels, Larry M. 1993. "Messages received: the political impact of media exposure." *American Political Science Review* 87(02):267–285.

Barthel, Michael, Elisa Shearer, Jeffrey Gottfried and Amy Mitchell. 2015. "The evolving role of news on Twitter and Facebook." *Pew Research Center* .

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

Benoit, Kenneth, Drew Conway, Benjamin E Lauderdale, Michael Laver and Slava Mikhaylov. 2016. "Crowd-sourced text analysis: reproducible and agile production of political data." *American Political Science Review* 110(2):XX–XX.

Berelson, Bernard. 1954. *Voting: A study of opinion formation in a presidential campaign*. University of Chicago Press.

Berger, Jonah. 2011. "Arousal increases social transmission of information." *Psychological Science* 22(7):891–893.

Berger, Jonah and Katherine L Milkman. 2012. "What makes online content viral?" *Journal of Marketing Research* 49(2):192–205.

Birkland, Thomas A. 1997. *After Disaster: Agenda Setting, Public Policy and Focusing Events*. Washington, D.C.: Georgetown University Press.

Bond, Robert M, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle and James H Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization." *Nature* 489(7415):295–298.

Bond, Robert and Solomon Messing. 2015. "Quantifying social media's political space: Estimating ideology from publicly revealed preferences on Facebook." *American Political Science Review* 109(01):62–78.

Bos, Linda, Sanne Kruikemeier and Claes de Vreese. 2016. "Nation Binding: How Public Service Broadcasting Mitigates Political Selective Exposure." *PLoS ONE* 11(5):1–11.

Brundidge, Jennifer. 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.

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

Bullock, John G. 2009. "Partisan bias and the Bayesian ideal in the study of public opinion." *The Journal of Politics* 71(03):1109–1124.

Centola, Damon and Michael Macy. 2007. "Complex contagions and the weakness of long ties." *American Journal of Sociology* 113(3):702–734.

Chadwick, Andrew. 2006. *Internet politics: States, citizens, and new communication technologies*. Oxford University Press.

Colleoni, Elanor, Alessandro Rozza and Adam Arvidsson. 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, Michael D., Bruno Gonçalves, Alessandro Flammini and Filippo Menczer. 2012. "Partisan Asymmetries in Online Political Activity." *EPJ Data Science* 1(1):1–19.

Dean, Jeffrey and Sanjay Ghemawat. 2008. "MapReduce: simplified data processing on large clusters." *Communications of the ACM* 51(1):107–113.

Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know About Politics and Why it Matters*. New Haven: Yale University Press.

DiMaggio, Paul, John Evans and Bethany Bryson. 1996. "Have American's Social Attitudes Become More Polarized?" *American Journal of Sociology* 102(3):690–755.

Donsbach, Wolfgang. 2004. "Psychology of news decisions factors behind journalists' professional behavior." *Journalism* 5(2):131–157.

Farrell, Henry. 2012. "The Internet's Consequences for Politics." *Annual Review of Political Science* 15:35–52.

Feezell, Jessica T. 2016. "Predicting Online Political Participation The Importance of Selection Bias and Selective Exposure in the Online Setting." *Political Research Quarterly* pp. XX–XX.

Fiorina, Morris P. and Samuel J. Abrams. 2008. "Political polarization in the American public." *Annu. Rev. Polit. Sci.* 11:563–588.

Flaxman, Seth R., Sharad Goel and Justin M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80(S1):298–320.

Galtung, Johan and Mari Holmboe Ruge. 1965. "The structure of foreign news the presentation of the Congo, Cuba and Cyprus Crises in four Norwegian newspapers." *Journal of Peace Research* 2(1):64–90.

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

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

Gentzkow, Matthew and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *The Quarterly Journal of Economics* (126):1799–1839.

Gerber, Alan and Donald Green. 1999. "Misperceptions About Perceptual Bias." *Annual Review of Political Science* 2(1):189–210.

Golbeck, Jennifer and Derek Hansen. 2014. "A method for computing political preference among Twitter followers." *Social Networks* 36:177–184.

Goode, Luke. 2009. "Social news, citizen journalism and democracy." *New Media & Society* 11(8):1287–1305.

Gottfried, Jeffrey and Elisa Shearer. 2016. *News Use Across Social Media Platforms 2016*. Pew Research Center.

Granovetter, Mark S. 1973. "The strength of weak ties." *American Journal of Sociology* pp. 1360–1380.

Groseclose, Tim and Jeffrey Milyo. 2005. "A measure of media bias." *The Quarterly Journal of Economics* pp. 1191–1237.

Guess, Andrew and Alexander Coppock. 2017. "Does Counter-Attitudinal Information Cause Backlash? Results from Three Large Survey Experiments.".

Guess, Andrew M. 2015. "Measure for Measure: An Experimental Test of Online Political Media Exposure." *Political Analysis* 23(1):59–75.
URL: *http://pan.oxfordjournals.org/content/early/2014/06/20/pan.mpu010.abstract*

Guess, Andrew M. 2016. "Media Choice and Moderation: Evidence from Online Tracking Data." Unpublished manuscript. Downloaded May 26, 2017 from https: //www.dropbox.com/s/uk005hhio3dysm8/GuessJMP.pdf?dl=0.

Hallin, Daniel C. and Paolo Mancini. 2004. *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman and James Franklin. 2005. "The elements of statistical learning: data mining, inference and prediction." *The Mathematical Intelligencer* 27(2):83–85.

Hindman, Matthew. 2008. *The myth of digital democracy*. Princeton University Press.

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, not ideology a social identity perspective on polarization." *Public Opinion Quarterly* 76(3):405–431.

Jamieson, Kathleen Hall and Joseph N Cappella. 2008. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press.

Kaplan, Andreas M. and Michael Haenlein. 2010. "Users of the world, unite! The challenges and opportunities of Social Media." *Business Horizons* 53(1):59–68.

Kenski, Kate and Natalie Jomini Stroud. 2006. "Connections between Internet use and political efficacy, knowledge, and participation." *Journal of Broadcasting & Electronic Media* 50(2):173–192.

Krosnick, Jon A. and Laura A. Brannon. 1993. "The impact of the Gulf War on the ingredients of presidential evaluations: multidimensional effects of political involvement." *American Political Science Review* 87(04):963–975.

Lazarsfeld, P.F., B. Berelson and H. Gaudet. 1944. *The peoples choice: How the voter makes up his mind in a presidential election*. New York: Duell, Sloan and Pearce.

Legewie, Joscha. 2013. "Terrorist Events and Attitudes toward Immigrants: A Natural Experiment." *American Journal of Sociology* 118(5):1199–1245.

Lelkes, Yphtach, Gaurav Sood and Shanto Iyengar. 2015. "The hostile audience: The effect of access to broadband Internet on partisan affect." *American Journal of Political Science* .

Levendusky, Matthew S. 2013. "Why do partisan media polarize viewers?" *American Journal of Political Science* 57(3):611–623.

Lewis, Seth C., Kelly Kaufhold and Dominic L. Lasorsa. 2010. "Thinking about citizen journalism: The philosophical and practical challenges of user-generated content for community newspapers." *Journalism Practice* 4(2):163–179.

Lippman, Walter. 1922. *Public opinion*. New York, NY: MacMillan.

McCombs, Maxwell E. and Donald L. Shaw. 1972. "The agenda-setting function of mass media." *Public Opinion Quarterly* 36(2):176–187.

Messing, Solomon and Sean J Westwood. 2012. "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online." *Communication Research* .

Munzert, Simon and Paul C. Bauer. 2013. "Political depolarization in German public opinion, 1980–2010." *Political Science Research and Methods* 1(01):67–89.

Mutz, D.C. 2002*a*. "Cross-cutting social networks: Testing democratic theory in practice." *American Political Science Review* 96(1):111–126.

Mutz, D.C. and L. Young. 2011. "Communication and Public Opinion: Plus Ça Change?" *Public Opinion Quarterly* 75(5):1018–1044.

Mutz, Diana C. 2006. *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Mutz, Diane. 2002*b*. "The Consequences of Cross-Cutting Circles for Political Participation." *American Journal of Political Science* 4:838–855.

Osborne, Miles, Saša Petrovic, Richard McCreadie, Craig Macdonald and Iadh Ounis. 2012. Bieber no more: First story detection using Twitter and Wikipedia. In *SIGIR 2012 Workshop on Time-aware Information Access*.

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

Prior, Markus. 2005. "News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout." *American Journal of Political Science* 49(3):577–592.

Prior, Markus. 2007. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge: Cambridge University Press.

Prior, Markus. 2009. "Improving Media Effects Research through Better Measurement of News Exposure." *The Journal of Politics* 71(3):893–908.

Prior, Markus. 2013*a*. "The Challenge of Measuring Media Exposure: Reply to Dilliplane, Goldman, and Mutz." *Political Communication* 30(4):620–634.

Prior, Markus. 2013*b*. "Media and Political Polarization." *Annual Review of Political Science* 16:101–127.

Puglisi, Riccardo and James M Snyder. 2015. "The balanced US press." *Journal of the European Economic Association* 13(2):240–264.

Rivers, Douglas. 2006. "Sample matching: Representative sampling from internet panels." *Polimetrix White Paper Series* .

Sears, David O and Jonathan L Freedman. 1967. "Selective exposure to information: A critical review." *Public Opinion Quarterly* 31(2):194–213.

Sinclair, Betsy. 2012. *The social citizen: Peer networks and political behavior*. University of Chicago Press.

Staab, Joachim Friedrich. 1990. "The role of news factors in news selection: A theoretical reconsideration." *European Journal of Communication* 5(4):423–443.

Stanford Mass Shootings of America project. 2016. Stanford Mass Shootings in America, courtesy of the Stanford Geospatial Center and Stanford Libraries. Technical report Stanford University. https://library.stanford.edu/projects/mass-shootings-america.

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

Sunstein, C. 2001. *Republic.com*. Princeton University Press.

Taber, Charles S. and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50(3):755–769.

Tewksbury, David. 2003. "What do Americans really want to know? Tracking the behavior of news readers on the Internet." *Journal of Communication* 53(4):694–710.

Tufekci, Zeynep. 2016. "The Real Bias Built In at Facebook." New York Times.
   URL: *http://www.nytimes.com/2016/05/19/opinion/the-real-bias-built-in-at-facebook.html*

Webster, James G. 2014. *The marketplace of attention: How audiences take shape in a digital age.* MIT Press.

Wegner, Daniel M. 1987. Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior*. Springer pp. 185–208.

Weng, Jianshu and Bu-Sung Lee. 2011. "Event Detection in Twitter." *ICWSM* 11:401–408.