AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets

Andrew M. Guess Princeton University

Abstract: *Does the internet facilitate selective exposure to politically congenial content? To answer this question, I introduce and validate large-N behavioral data on Americans' online media consumption in both 2015 and 2016. I then construct a simple measure of media diet slant and use machine classification to identify individual articles related to news about politics. I find that most people across the political spectrum have relatively moderate media diets, about a quarter of which consist of mainstream news websites and portals. Quantifying the similarity of Democrats' and Republicans' media diets, I find nearly 65% overlap in the two groups' distributions in 2015 and roughly 50% in 2016. An exception to this picture is a small group of partisans who drive a disproportionate amount of traffic to ideologically slanted websites. If online "echo chambers" exist, they are a reality for relatively few people who may nonetheless exert disproportionate influence and visibility.*

Verification Materials: The materials required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/ZFE3NE.

When it comes to political content, people choose what they like. This simple claim forms the core of more than a half-century of social science research on selectivity in exposure to news about politics. Uniting this broad array of scholarship is a longstanding concern: that people, always seeking reinforcement, prefer to wall themselves off from contradictory sources of information. This has remained a constant even as the underlying worries change with the times, from the influence of propaganda (Berelson, Lazarsfeld, and McPhee 1954; Klapper 1960; Lazarsfeld, Berelson, and Gaudet 1944), to the dominance of commercial imperatives (Katz 1996), to the rise of cable television (Arceneaux and Johnson 2013; Jamieson and Cappella 2008; Prior 2007), the spread of the internet (Garrett 2009; Hindman 2008; Valentino et al. 2009), and most recently, the ubiquity of online social media (Bakshy, Messing, and Adamic 2015; Messing and Westwood 2014).

New technological developments have a tendency to refocus attention on the age-old question of democratic competence: To what extent do citizens meet the basic requirements for informed collective decision making? Theorists and political scientists alike, especially within the deliberative tradition, hold that a critical precondition is exposure to competing viewpoints (Downs 1957; Fishkin 1991; Habermas 1984; Mutz 2006; Shapiro 2013). Others further emphasize the importance of shared experiences within a polity (e.g., Sunstein 2017). Both conditions implicate the quality and availability of information about public affairs.

Andrew M. Guess, Assistant Professor of Politics and Public Affairs, Fisher Hall, Princeton University, Princeton, NJ 08544 (aguess@princeton.edu).

This study was approved by the New York University Institutional Review Board (IRB-FY2016-1342). I would like to thank the editors and three anonymous reviewers for their detailed guidance and feedback on this article. I am grateful to Pablo Barberá, Neal Beck, Noah Buckley, Alex Coppock, Pat Egan, Albert Fang, Don Green, Trish Kirkland, Jeff Lax, Lucas Leemann, Yph Lelkes, Jonathan Nagler, Brendan Nyhan, Markus Prior, Jason Reifler, Robert Shapiro, Gaurav Sood, Lauren Young, and seminar participants at the Columbia University Department of Political Science, the Annenberg School for Communication at the University of Pennsylvania, the NYU Center for Data Science, and the Yale ISPS Experiments Workshop for extremely helpful comments and suggestions. Thanks also to those who provided valuable feedback during seminars at Brown University, Princeton University, Rutgers, Penn State, and NYU Abu Dhabi. I additionally benefited from comments by discussants and attendees at the 2016 Southern Political Science Association and Midwest Political Science Association annual meetings and the 2016 APSA Political Communication Pre-conference at Temple University. I am indebted to Doug Rivers, Brian Law, and Joe Williams at YouGov for facilitating access to the 2015 Pulse data, and to Ashley Grosse for making possible the survey on privacy attitudes. The 2016 data collection was generously supported by the American Press Institute. Some of the analysis was made possible by High Performance Computing (HPC) clusters at New York University.

American Journal of Political Science, Vol. 65, No. 4, October 2021, Pp. 1007–1022

 DOI: 10.1111/ajps.12589

Today, people increasingly find (or avoid) such information online. According to the Pew Research Center, in 2019, 82% of Americans said they get news online “at least sometimes,” and over half (57%) do so “often” on a mobile device (Walker 2019). Although online media are not yet the primary source of political information for all Americans, they are for the youngest age groups, and the upward trend continues. As websites, social media platforms, and mobile devices become the dominant mode of news delivery, the range of available sources continues to multiply. Early observers of the internet’s impact on society saw this proliferation of choice as fundamental to its transformative power, but they split between near-utopian visions of a “global village” without boundaries and gloomy predictions of “cyber-balkanization” (Benkler 2006; Negroponte 1995; Van Alstyne and Brynjolfsson 2005).

The pessimistic view is now far more prevalent. In the worst-case scenario, people elect to consume only ideologically congenial information, resulting in an “echo-chamber” effect and, ultimately, increasing polarization. Additionally, some worry that hidden algorithms could speed along this process by replicating and reinforcing people’s preferences on social media without their awareness (Pariser 2011). This is exemplified by frequent lamentations in the popular press about the ideological cocooning of America via social media (e.g., Klasa 2016; Madrigal 2017). These concerns evidently resonate with the mass public: People generally believe that consulting like-minded sources for election news is common, especially among their political opponents (Perryman 2017).

Employing a combination of individual-level web and mobile consumption data from 2015 and 2016, surveys, and supervised text classification of millions of web pages, I demonstrate that this widely shared characterization is exaggerated. Instead, the picture is chiefly one of moderation: On average, the media diets of Democrats and Republicans overlap more than they diverge and center around the middle of the ideological spectrum. This pattern appears to be at least partially driven by the dominance of mainstream, relatively centrist websites such as CNN.com and MSN.com. Most strikingly, the media diets of the vast majority of people—regardless of political orientation—are reasonably moderate, with only a small share of highly partisan respondents driving a disproportionate amount of traffic to the most ideologically extreme outlets, especially on the right (see Jamieson and Cappella 2008). Media consumption habits are remarkably comparable across the political spectrum, and although there is a correlation between individuals’ partisan or ideological predispositions and the overall political slant of their media diets, this correlation is quite modest.

This article proceeds as follows. First, I discuss the potential role of informational intermediaries such as web portals in structuring people’s online news habits. Next, I briefly discuss how my research approach addresses many of the well-known methodological difficulties of studying partisan selective exposure. In the following section, I outline the data sources, the steps I take to process and validate them, and the measures I use to summarize and compare individuals’ political news consumption patterns. In presenting the results, I begin by focusing on respondents’ media diets and the factors associated with their relative ideological lean. I also show that, because people with more slanted diets also tend to consume more partisan news overall, these individual-level patterns can still aggregate to produce a seemingly polarized political news audience—one with outsize impact on the political process, as I demonstrate with verified voting data. A discussion then concludes with directions for future research.

Portals as Information Intermediaries

Aside from measuring the extent of online selective exposure, it is important to understand the mechanisms through which people encounter politically relevant news. Web visit data allow me to focus on a set of mechanisms related to structural features of the online browsing experience, specifically default settings such as homepages and bookmarks that, in the language of behavioral science, constitute a kind of digital choice architecture for users (Lorenz-Spreen et al. 2020; Sunstein and Thaler 2008). Effectively, individuals can be “nudged” to use highly trafficked web portals and news aggregators—“general-interest intermediaries” that facilitate a kind of virtual public sphere. For Sunstein (2017), these intermediaries should provide two core preconditions for a deliberative republic: shared experiences and the possibility of serendipitous encounters with diverse perspectives. It is not hard to see how large informational hubs can serve these purposes, both by offering easily accessible news about current affairs with a relatively moderate slant and by exposing visitors to unexpected material from different realms (e.g., world news, entertainment, and sports). If many people approach their daily informational needs by satisficing in this way, features such as lists of trending topics and the most popular articles of the day could help to magnify these patterns of shared exposure and serendipitous discovery. The cumulative effect of this form of passive information consumption would be a tendency toward balance or moderation in people's media diets. On the other hand, social media has enabled passive consumption of an altogether different kind, fueling concerns about predictive personalization that subtly arranges content according to revealed preferences (Pariser 2011). According to this account, social feeds can gradually become more ideologically homogeneous and extreme as people engage with the tailored information they encounter, generating a feedback loop of polarization (see Shmargad and Klar 2020). But this prediction arguably rests on a critical ingredient: a nonnegligible amount of political content available to be ranked according to a user's preferences, which itself requires a level and type of political interest shared by the user and her larger network of social connections. For many people, these preconditions might not hold and even this form of passive (social) consumption would therefore not lead to greater levels of selectivity (e.g., Eady et al. 2019). To test the above predictions, I will explore the relationship between polarized information consumption (or lack thereof) and the prevalence of news portals, relatively centrist mainstream media sources, and political content in people's news diets.

Methodological Challenges

Since the critical review of Freedman and Sears (1965), skeptics have questioned whether people are as selective in their choices of how to receive information as previously supposed. More recently, advances in measurement and data collection have backed up some of these claims. For instance, Gentzkow and Shapiro (2011) use web traffic data to conclude that online ideological segregation—an aggregate measure of political isolation in news consumption patterns—is less severe than for national newspapers or in face-to-face social networks. On Twitter, Barberá (2014) uses panel data to show evidence of cross-cutting follow patterns that lead to moderation rather than polarization. Using data on the web browsing histories of a sample of Bing Toolbar users, Flaxman, Goel, and Rao (2016) find that social media and search engines can drive ideological diversity in exposure to online news. Likewise, Scharkow et al. (2020) use random-effects within–between models to show that social media and web portals can lead to increased visits to news sites.

This body of work sits alongside research uncovering some evidence of selective exposure in relatively higher-choice media environments. Using a simulated online environment, Garrett (2009) finds selectivity among subjects recruited from partisan websites but finds more of a preference for reinforcing content than an aversion to challenging information. A number of other innovative studies incorporate unobtrusive measurement into designs that focus subjects' attention on seemingly unrelated tasks, such as clicking through a CD-ROM with candidate issue positions (Iyengar et al. 2008) or browsing through magazines in a waiting room (Stroud 2011), to test for correlations with political predispositions or other preferences over content. All of these effects are relatively modest. Other studies experimentally manipulate source labels in order to test for revealed preferences (e.g., Iyengar and Hahn 2009). What unites these designs is that they rely on a predefined set of media choices to ensure internal validity. The costs of this tradeoff are increasingly evident as the real-world choice set becomes, in principle, limitless.

Another challenge inherent to the study of selective exposure is simply that researchers are aiming at a moving target: The pathways by which people encounter information online evolve quickly, such as the shift from visiting homepages to relying on links from social media.[1] But even within a given medium—such as the focus of this study, news sources on the web accessed from laptops, desktop computers, and mobile devices—there are mixed results in the literature. An important reason is that errors in survey-based measures of exposure to media content are likely correlated with political identities, which could in turn exaggerate evidence for selectivity. As Prior (2013) argues, "taking self-reports at face value requires the assumption that the very people who follow their wishful thinking when they evaluate economic performance or perceive centrist news as biased faithfully report when they turn off the pro-attitudinal message stream or follow counter-attitudinal programming." Relatedly, in previous work, I show that the most commonly used survey-based self-report measures are often inaccurate in the context of online political media (Guess 2015). Both studies instead recommend the use of direct, behavioral measures of media exposure via passive metering technologies.

Given the limitations of survey-based approaches to studying media exposure, I follow the lead of pioneering scholars who turned to third-party data on web visits

[1]See http://www.poynter.org/2014/3-takeaways-from-the-death-of-the-homepage-and-the-new-york-times-innovation-report/252632/.

to better understand online media consumption patterns (Gentzkow and Shapiro 2011; Tewksbury 2003). Their work generated useful macrolevel evidence by aggregating to the level of individual web domains. Research using data from providers such as comScore has continued to provide rich insights into the composition of the online news audience (Nelson and Webster 2017).

My approach is most related to Dvir-Gvirsman, Tzfati, and Menchen-Trevino (2014), who hand-coded web tracking data from a sample of Israeli internet users in an online survey. What sets the present study apart from most existing research on the topic of online selective exposure is that it combines individual survey responses with digital trace data (see Stier et al. 2019 for an overview of this emerging approach). In other words, I follow traditional political science research in beginning with surveys on a well-defined population offering a rich array of covariates on respondents' demographic and political attributes. At the same time, I obtain passive metering data covering visits to online news articles that are linked to each respondent rather than aggregated to broad categories. This means that problems of ecological inference are not a concern as they are with studies lacking the survey component (e.g., Flaxman, Goel, and Rao 2016; Gentzkow and Shapiro 2011). Using browsing data as opposed to site-level audience metrics also means that I can more directly test expectations related to selective exposure, which is typically conceptualized as an individual-level behavior (Stroud 2008). Finally, because this study tracks respondents in their real-world environments, there is no artificial restriction of the choice set; I use a measure of site-level ideological slant, detailed next, that covers nearly 500 web domains.

Data

My primary data sources are two national surveys merged with panelists' internet browsing histories. These data were collected by the online polling firm YouGov in conjunction with the passive metering service Wakoopa. Aside from financial transactions and passwords, there are no restrictions on the types of websites that can be included in the data. Moreover, unlike the approach used in Guess (2015), the software tracks web traffic for *all* browsers installed on a user's computer. The online tracking panel is currently branded as YouGov Pulse, and panelists are recruited from YouGov's traditional participant pool via incentives. Those who installed the Wakoopa software on their computers provided explicit consent for sending anonymized web visit data to YouGov, which in turn agreed not to share it with third-party vendors. Users can pause the metering for 15-minute increments and are free to uninstall and leave the panel at any time.

I separately analyze data from early 2015 and from October 2016, shortly before the U.S. presidential election. The 2015 data set contains 6,319,441 observations at the respondent-visit level, covering panelists who installed the passive metering software on their desktop or laptop computers (excluding mobile phones). This sample includes full URLs of page visits over a 3-week period in 2015, from February 27 to March 19. Because respondents ($N = 1,392$) were not recruited using probability sampling, I construct sample weights by raking to population marginal distributions of age, gender, race, region (all from 2014 Census estimates), and party identification (from 2014 Pew Research Center estimates). The 2016 data were collected for a study on the consumption of online misinformation and cover 16,984,969 total visits ($N = 2,512$ respondents) from October 7 to 31, 2016 (Guess, Nyhan, and Reifler 2020). When analyzing this data set, I use the supplied sample-matching weights.[2] Finally, for a subset of 2016 respondents ($N = 654$), I analyze domain-level mobile visit data covering the same time period ($491,983$ visits). Table A1 in the Online Appendix summarizes the demographic characteristics of the two main samples before weighting. The raw samples skew younger and somewhat more Democratic than the general population, although the gender and education breakdowns are fairly representative.

Given the opt-in nature of the Pulse panel and the mode of data collection, it is important to establish the validity of the sample by examining potential biases in its composition. As mentioned, the use of raking weights (for 2015) and sample-matching weights (for 2016) in all analyses helps to address potential selection biases on observable characteristics.[3] Still, there is a remaining possibility that the additional step of installing passive metering software, however unobtrusive, could generate systematic differences between YouGov respondents who opt in to Pulse and those who do not. Although I cannot investigate all such possibilities, it would be particularly worrisome if attitudes theoretically related to Pulse data collection were predictive of both the likelihood of joining the panel as well as other politically relevant characteristics. To test whether this could be the case, I show in Online Appendix B that responses to survey questions

[2] I use raking weights for 2015 because sample-matching weights were not available for the data collected up to that point.

[3] See Online Appendix B for details on YouGov's sampling methodology.

about online privacy are reassuringly comparable between those who opt in to Pulse and those who do not.

Methods

Measuring Media Slant

Data linking individuals to media consumption provide the heart of the evidence in this analysis. However, an additional component is necessary: a valid method of scaling the ideological slant of online media sources. The limitation shared by existing methods for the purposes of this study is the relatively small number of websites and publications with ideological slant scores. Perhaps the most well-known method, that of Groseclose and Milyo (2005), covers 20 newspapers and television shows. A more recent content-based attempt, by Budak, Goel, and Rao (2016), covers 15 sources. Gentzkow and Shapiro (2010) estimate scores for many more publications—435—but all are print newspapers.

To solve this problem, I make use of "alignment" scores produced by Bakshy, Messing, and Adamic (2015) to study exposure to news and opinion content shared on Facebook. Fundamentally, the method is extremely simple: A web domain's alignment score is the average of the self-reported ideology of users who have shared pages from that domain on Facebook. By essentially leveraging the crowdsourced judgments of millions of Facebook users, this approach avoids relying on coding schemes or similarity to particular phrases. And by using data on shares, it captures behavior that is both meaningful and distinct from consumption. This is critical because it avoids circularity that could otherwise arise if, for example, web visit or click data from known ideological groups were used as a proxy for media slant to then test against similar web visit data. Furthermore, less controversially than private consumption behavior, sharing content on social platforms plays a well-known role in helping people maintain their self-image among peers in relevant social groups, including partisan and ideological groups (Bowyer and Kahne 2019; Weeks et al. 2017). As elaborated in Online Appendix C, this suggests that measurement error would likely cut against a finding of balance and overlap in people's online political news diets. Scores range from approximately -1 to 1 and, reassuringly, correlate highly with existing estimates of media slant.[4]

[4]Additional information, including graphical and numerical representations of domain alignment scores, is available in Online Appendix C.

This procedure generates valid proxies of media outlet ideology for *relative* comparisons—that is, the ability to say that a source is to the left or right of another on an implied ideological continuum. Like other prominent attempts to estimate media slant, this approach does not allow reliable identification of the *absolute* ideological position of any outlet. This means that the midpoint of the scale used in this study (i.e., 0) is arbitrary and should not be interpreted as corresponding to an objective benchmark of ideological balance (or a lack of ideological slant). Neither the measurement strategy nor the analyses in this study solve the fundamental unobservability problem identified by Groeling (2013): We cannot know the objective characteristics underlying media coverage without relying on the media itself, so that any notion of unbiasedness is inherently subject to contestation. Thus, as Gentzkow and Shapiro (2010) note in their study of newspaper ideology, such estimates should not be interpreted as "a benchmark of 'true' or 'unbiased' reporting" (pp. 36–37).

Classifying Political News at the Article Level

To separate pages related to news about politics from the rest, I use a penalized logistic regression model trained on text features of scraped articles with labels derived from section headers embedded within URLs (Flaxman, Goel, and Rao 2016). The motivation for this approach is that regularization is necessary to tame models with large numbers of correlated predictors: Certain words or phrases may tend to co-occur in news articles, for instance, which would lead to severe multicollinearity. Regularization shrinks the magnitude of estimated coefficients, which could otherwise grow arbitrarily large. This reduces overfitting and improves classification accuracy. Overall, performance is extremely high: The 2015 model has a within-training-set accuracy of 89.7% and recall of 88.9%, and the 2016 model has an accuracy of 90.6% and recall of 88.6%.

I use the classifiers to dichotomize each relevant page URL as related to news about politics or not based on a cross-validated threshold. For both 2015 and 2016, this means that about one-fifth the total number of visits to pages with associated alignment scores are estimated to be about politics. I provide details on these results and the procedures I use for article scraping, estimation, and validation in Online Appendix D.

Measures and Empirical Approach

Finally, to facilitate analysis at the individual level, I produce a measure of the overall ideological lean of each respondent's media diet by simply averaging over the alignment scores for the domains associated with his or her web visits in the Pulse data. For example, if a respondent made 10 visits to pages from a left-leaning site at −0.15 and one visit to a right-leaning site at 0.10, her average would be −0.127.

Given the amount of researcher discretion involved at each stage, I apply different combinations of preprocessing and analytic steps to the data and report the results of each either in the article or the Supporting Information. The primary decisions are as follows:

- Whether to apply the political news classifier: I either use the raw data, counting all visits to pages whose domains have an associated alignment score, or I subset the data to political news only according to the classifiers.
- Whether to apply sample weights: In plotting the distribution of the average ideological slant of individuals' media diets, I either use or ignore weights that adjust for selection into the sample.
- Whether to filter out repeated visits to the same page: A reasonable preprocessing decision is to remove sequential duplicates, that is, additional visits to a given page by the same person, one after the other. This accounts for automatic refreshes in open tabs and other aberrations in web visit patterns.[5] When applied, the deduplication process reduces the number of URLs in the data by 59.2% for 2015 and by 20.1% for 2016.
- Whether to weight individual-level media diets by the length of time spent at each page rather than the number of visits: Pulse data include duration information for each page visit, although the reliability of this measure is untested. Regardless, I will show that my primary measure of media diet slant is robust to weighting by duration rather than the number of visits.
- Whether to apply domain exclusion rules: In all main analyses (which I refer to as the baseline), I exclude visits from twitter.com, youtube.com, and amazon.com, which have associated alignment scores despite being platforms or, in the latter case, an online retailer.[6] I drop visits to pages on the twitter.com and youtube.com domains because I lack measures of ideological slant for individual pieces of content shared on those platforms, which could greatly differ from each other. Amazon.com is not specifically relevant to the substance of my analysis, so I exclude it as well. Aside from this processing decision, I explore dropping visits to web portals such as MSN and AOL, which as hypothesized may have a moderating role in people's media consumption.

To answer questions about the extent of selective exposure in online media consumption, I explore the distribution of respondents' average media diets broken down by party.[7] I first do this graphically, then report summary statistics. In addition to medians, means, and standard deviations, I compute a simple statistic, the "overlapping coefficient" (Clemons and Bradley Jr. 2000; Inman and Bradley Jr. 1989), which is the shared area underneath two probability density curves as a fraction of the total area. It ranges from 0, which indicates completely disjoint distributions, or no overlap, to 1, which means that the two distributions are identical. (See Lelkes 2016 for a recent application in political science.) The formula for this coefficient is

$$1 - \frac{1}{2}\int_{-\infty}^{+\infty} |f(x) - g(x)|dx. \qquad (1)$$

The machinery outlined thus far is designed to provide simple and intuitive metrics for quantifying the political orientation of people's online news consumption habits, as well as for facilitating comparisons across subgroups and time. Although powerful, they do not in themselves suggest the most appropriate interpretation of results. The literature on partisan selective exposure is largely silent on the question of how much of a preference for congenial content is acceptable or desirable, in part because finding a significant individual-level correlation of any magnitude between political predispositions and media choice is sufficient to reject the null hypothesis. This approach does not readily translate to studies examining the *distribution* of media diets within the population and the *extent* to which they reflect politically homogeneous consumption patterns, which draw on but cannot completely adjudicate between competing theoretical or normative accounts.

Returning to the overlap measure, should any pattern short of complete correspondence between two

[5]Before removing sequential duplicates, I also drop all anchors from URLs—the "#" followed by text pointing to a specific part of a given page. This ensures that the deduplication process captures visits to multiple sections within the same page.

[6]In Online Appendix A, I report results without dropping pages from these domains.

[7]In all analyses, leaners are not coded as partisans.

groups' information environments—that is, an overlapping coefficient of 1—be described as an echo chamber? Such a stringent standard is not a faithful representation of the concerns of scholars such as Sunstein, and it likely conflicts with other societal interests, such as providing space for enclaves to flourish and allowing people to develop their own diverse standards of information quality. At the same time, having no overlap at all—a coefficient of 0—would clearly trigger concerns about shared reference points and the ability for members of the public to deliberate meaningfully. To the extent that this metric can inform the debate, a reasonable baseline should therefore fall between its two extremes. But theory does not offer clear guidance as to where that should be. How much cross-cutting information people *should* encounter is an open and likely irresolvable question. Defining as sufficient any amount greater than zero virtually preordains a rosy conclusion, yet requiring of citizens that they continuously engage with challenges to their worldviews fundamentally neglects their autonomy.

For the time being, then, research on these questions must progress by balancing atop this theoretical seesaw. Because it will nonetheless be useful to have a consistent reference point throughout the ensuing discussion, I choose the exact midpoint (0.5) as a tentative standard for the minimum degree of overlap between partisan groups' media diets needed to sustain an interpretation of meaningfully shared informational environments. To avoid teetering into unsupported assertions on either side, I strive to convey nuance and uncertainty as well as the possibility of multiple interpretations of the evidence that I present below.

Results

I focus on two versions of the results for each year's data. Within each panel of Figure 1, I separately plot the sample-weighted densities of the average ideological lean of self-reported Democrats', Republicans', and independents' media diets incorporating only visits to pages selected as political news according to the supervised classifier. The second panel of each row shows the density of visits to articles about news and politics after first dropping pages from "portal" domains.[8] For independents, Republicans, and Democrats, most respondents' average media diets are close to the midpoint of the ideological spectrum. The peaks of the density curves all fall slightly to the left of center, reflecting the somewhat left-of-center estimated lean of most prominent mainstream news sources reflected in the alignment scores.[9]

This does not suggest that the media diets of Democrats and Republicans perfectly overlap, of course—more Democrats have somewhat more left-leaning news consumption tendencies, as evidenced by greater probability mass on the left half of the distributions. This pattern is likely driven by outlets such as NYTimes.com (−0.55) and BuzzFeed (−0.52) whose alignment scores lean fairly liberal. This does not necessarily reflect an inherent ideological bias in the content of these sources' news product; as noted above, alignment scores provide informative proxies of relative, not absolute, slant. The curve for Republicans is not a mirror image of this pattern; if anything, the density is highly symmetrical around the center with the exception of a cluster (about 10%) of those respondents whose media consumption is concentrated on the farther right end of the ideology scale.

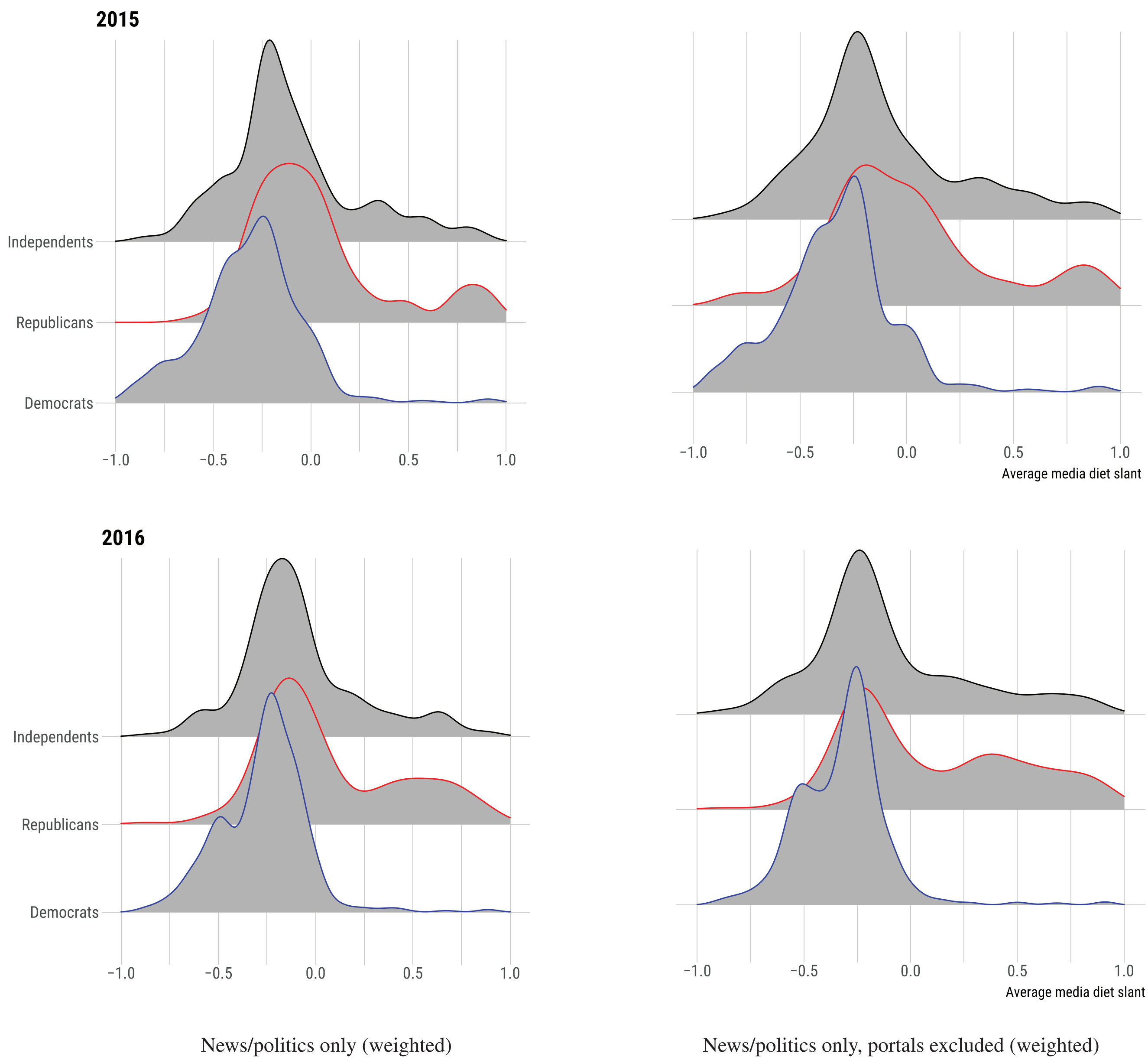
Putting some numbers to these graphical impressions, the top half of Table 1 characterizes the data shown in the top of Figure 1. Focusing on the top sample, the median slant of all respondents' media diets is −0.213 and the mean is −0.135—somewhat left of center out of a scale ranging from −1 to 1. Moving across the first row of the table, it is clear that the median for Democrats is to the left of the median for Republicans. The median for Republicans is actually slightly to the left of center, and the mean is slightly to the right. The standard deviation for Republicans (0.35) is larger than that for Democrats (0.262), likely driven by the presence of some respondents farther to the right with more homogeneously conservative media diets in addition to those closer to the center. But this latter and much larger group ensures a substantial overlap in the distributions for Democrats and Republicans: The overlapping coefficient is 0.63, implying that almost 65% of the total distribution plausibly describes the average media diet slant of typical Democrats or Republicans. Regardless of how the data are processed and analyzed, the overlapping coefficient ranges from 0.61 to 0.74, well above the tentative baseline of 0.5 suggested in the previous section.

In sum, the 2015 results show some moderation in respondents' overall media diets regardless of partisan affiliation. Whether Democrat, Republican, or

[8] www.aol.com, www.msn.com, www.google.com, and www.yahoo.com. As with AOL, MSN, and Yahoo! (but not Yahoo! News), I treat Google as an all-purpose informational hub. See Online Appendix A for results with different processing steps.

[9] The median of the 495 available alignment scores is −0.059. See Online Appendix G for details on how I categorize "mainstream" sites.

FIGURE 1 Americans' Online Media Diets by Partisanship



Note: Top: Distribution of the slant of respondents' online media diets, Feb. 27-March 19, 2015. Bottom: Distribution of online media diet slant, Oct. 7-31, 2016.

independent, the large bulk of these individuals' media diets cluster around the center of the ideological spectrum. Another way of showing this empirically is that more than 66% of Republicans in the sample have average media diets between −0.25 and 0.25 (one-quarter of the range); the corresponding figures for Democrats and independents are 51% and 55%. Furthermore, as Figure A6 in the Supporting Information shows, this moderation is mainly a reflection of the centrism of domains visited by most respondents, not a balancing of sources with opposing slant, although there are a small number of individuals with media diets close to 0 who appear to sample over a more ideologically diverse set of sources.[10] There is a visible correlation between party and the average lean of the media diets—the blue curve, representing Democrats, is somewhat "fatter" to the left of center, reflecting the difference in the medians for Democrats and

[10]More precisely, there are six respondents (0.6%) with online media diets whose average absolute slant is less than 0.5 but have variance greater than 0.5.

TABLE 1 Americans' Online Media Diets: Summary Statistics by Partisanship

	All			Democrats			Republicans		
2015									
News/politics only (weighted)	−0.213	−0.135	(0.346)	−0.296	−0.312	(0.262)	−0.051	0.027	(0.350)
News/politics only, portals excluded (weighted)	−0.214	−0.142	(0.371)	−0.309	−0.329	(0.268)	−0.047	0.019	(0.399)
2016									
News/politics only (weighted)	−0.165	−0.106	(0.332)	−0.245	−0.272	(0.219)	−0.060	0.078	(0.366)
News/politics only, portals excluded (weighted)	−0.218	−0.121	(0.376)	−0.293	−0.322	(0.213)	−0.018	0.105	(0.412)

Note: Median, mean, and s.d. (in parentheses) corresponding to data in Figure 1.

Republicans shown in Table 1—but, as discussed below, it is modest.

The same basic patterns are evident for 2016 in the second row of Figure 1: There is a significant degree of overlap between the distributions for Democrats and Republicans, with both centered close to the middle, and a group of Republican respondents whose media diets are much more conservative than those of the rest of the sample. A visual inspection of the plots suggests more of a "cocoon" on the right than was evident in 2015.[11] The second-to-last column of Table 1's bottom half shows a clear rightward shift in the mean from 2015, which when compared to the medians suggests a right skew in the data. This, along with the larger standard deviations for the Republican distributions, is consistent with the Republican group containing both people with relatively moderate media consumption habits and a smaller subset (close to 30%) with consistently conservative media diets. Due to the greater skew among Republicans in 2016, the overlapping coefficient with Democrats falls under the midway point (0.46), although using different processing steps (featured in Online Appendix A), the overlap reaches up to 0.60.

The left and right panels in each row of Figure 1 suggest that portals play a subtle role in people's online media diets. In terms of overlap, the distributions are similar: In 2015, the overlapping coefficient is virtually unchanged when portals are dropped from the slant estimates, whereas in 2016 the overlap decreases slightly from 0.46 to 0.41. Measures of central tendency indicate, on the other hand, that inclusion of portals serves to anchor the distributions closer to the center of the ideological spectrum. As shown in Tables A2 and A3 of Online Appendix A, the median estimated slant shifts markedly to the left overall and for Democrats in both 2015 and 2016 when portals are dropped, and the standard deviation is larger for all groups. This illustrates that the degree of overlap and relative moderation seen in respondents' information diets is not an artifact of the decision to include visits to popular news portals when computing individual-level summaries. At the same time, by some measures, portals are associated with pulling news consumption further toward the center—suggestive evidence of the importance of informational hubs in limiting the extent to which people cocoon themselves into ideologically congenial news diets. I provide additional evidence that both portals *and, as suggested above, major mainstream news sites in general* are significant drivers of this relatively moderate pattern in Online Appendices F and G.

Clearly, the evidence in 2015 is more consistent with the hypothesis that the internet facilitates exposure to cross-cutting information than what the data collected 1.5 years later, during the throes of a highly contentious presidential election campaign, indicate (Peterson, Goel, and Iyengar 2019). Whereas the overlapping coefficients calculated for 2015 are all well above the somewhat arbitrary 0.5 threshold I identified as a baseline, the analogous statistics for 2016 do not always meet this standard. Additionally, estimates of the share of Republicans (51%), Democrats (39%), and independents (48%) with media diets between −0.25 and 0.25 are smaller than the comparable figures for 2015. The evidence is thus consistent with arguments that the online media landscape is continuing to polarize, although it is important to stress that because the two data sets were collected at different points in each year, the differences could also reflect seasonality or other idiosyncrasies in the intervening period.

[11]This is most obvious when sample weights are *not* applied; see Figures A3 and A4 in the Supporting Information.

In Figure A8 of Online Appendix A, I show that the above patterns are largely replicated for the subset of respondents for whom mobile data were available during the same period in 2016. The mobile data provide domain-level information only, so classification of pages is not possible. Still, given speculation that the evolution of news consumption practices might be a driver of online selective exposure, it is striking to see broad overlap—a coefficient of 0.59, beyond the suggested 0.5 threshold—of the distributions of media diet slant in an environment in which distraction and susceptibility to partisan shortcuts might be more pronounced.[12]

Predictors of Media Diet Slant

Next, I explore the determinants of the overall slant of respondents' media diets. In Table 2, I show the results of simple linear regressions of media slant on individual-level characteristics using the 2015 and 2016 data:[13]

$$Y_i = \alpha + \beta_1 \text{Democrat}_i + \beta_2 \text{Republican}_i + \beta_3 \text{Independent}_i + \gamma \mathbf{X}_i + \epsilon_i, \qquad (2)$$

where $\mathbf{X}_i$ is a vector of demographic characteristics that includes age (four categories; reference category: 18–29), race (four categories; reference category: "Other"), gender, family income (12-point scale), and educational attainment (five categories; reference category: less than high school). (The reference category for party identification is "Other/don't know.") The media slant measure used as the dependent variable corresponds to the left panels of Figure 1, which include only pages classified via the supervised learning procedure as related to political news. Both models show a partial correlation between party identification and the overall ideological slant of one's media diet. In 2015, being a Democrat was associated with having a more liberal media diet, pushing respondents' overall slant leftwards by nearly 13% of the full range of the ideological spectrum on average. Additionally, even after controlling for party identification, all age categories above 29 are associated with more conservative media diets (for related evidence, see Guess, Nagler, and Tucker 2019). The results for 2016 are broadly similar, with the exception that the coefficient for Democrats is somewhat smaller and the coefficient for Republicans is also significant, associated with a more conservative media diet by about 6% of the range of the ideological spectrum. These results show that there is in fact an association between partisan affiliation and media consumption. However, this association is reasonably modest.

The Skewed Audience for Political News

The results so far have focused on summary measures intended to capture the overall ideological lean of individuals's news media consumption patterns. This approach reveals a relatively moderate portrait of most respondents' media diets with a large degree of overlap between Democrats and Republicans. Given widespread concern that consumption patterns are much more skewed than they appear to be—namely, belief in a nonoverlapping distribution of media exposure in which Democrats and Republicans encounter only like-minded content online—it is worth delving into the website activity underlying the numbers discussed so far. What does the audience for political news look like from a macro perspective?

Figure 2 plots the distribution of news visit slant, with observations pooled across respondents and grouped by the partisanship of the originating user. Here, we see something closer to the popular image of online ideological cocoons, most identifiably a large and distinct "echo chamber" of conservative media sources on the right side of the spectrum. But even here the data suggest a somewhat different story: There is another similar-sized peak among Republicans centering roughly around the middle, overlapping with a corresponding peak for Democrats. A comparison to Figure 1 suggests that traffic to the most partisan websites is actually coming from a small proportion of respondents (which can be seen more directly in Figure A7 in the Supporting Information). To illustrate, the 8% of Republican political news consumers in 2015 with media diets whose average slant is estimated to be greater than 0.75 (roughly, outlets at and to the right of Fox News) drive more than 24% of all political news visits among that group. Based on their media consumption patterns, these individuals are not representative of most others in their group. For instance, Figure A5 in the Supporting Information shows that even many people who identify in the data as very conservative have relatively moderate media preferences. But those who do not are driving a disproportionate amount of traffic to conservative sites, producing at the macro level an illusion of polarized media consumption.

This evidence, then, is consistent with a view that—among the fraction of respondents who visit news and

[12]Aside from potential sample attrition, another important limitation of the mobile analysis is a lack of app use data.

[13]Online Appendix E presents models that additionally include ideological self-placement.

TABLE 2 Determinants of Media Diet Slant

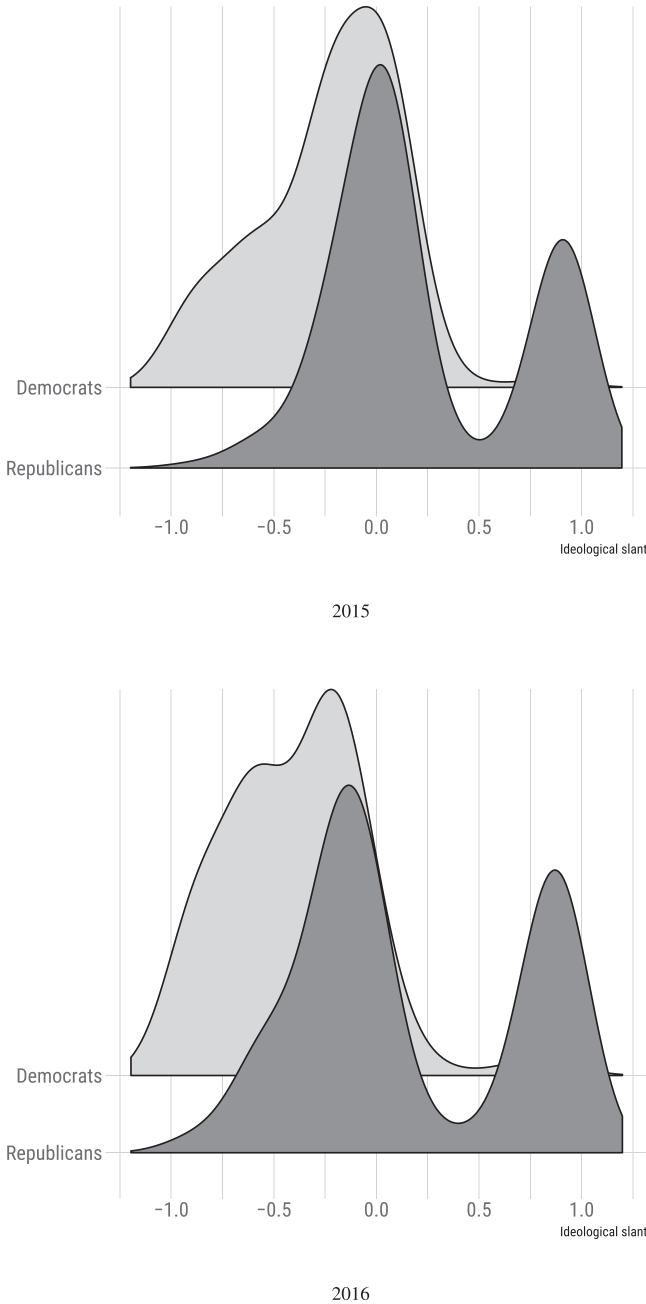
	Average media diet slant (news/politics only)	
	2015	**2016**
30–44	0.136**	0.062
	(0.034)	(0.035)
45–59	0.205**	0.123**
	(0.037)	(0.029)
60+	0.242**	0.203**
	(0.051)	(0.028)
White	−0.038	0.021
	(0.037)	(0.040)
Black	0.010	−0.047
	(0.047)	(0.046)
Hispanic	−0.017	0.110
	(0.054)	(0.066)
Female	−0.022	−0.024
	(0.030)	(0.021)
Income level	−0.002	0.005
	(0.005)	(0.003)
High school	0.019	0.079
	(0.071)	(0.073)
Some college	−0.048	0.055
	(0.062)	(0.071)
College graduate	−0.026	0.017
	(0.068)	(0.071)
Postgraduate	−0.117	−0.006
	(0.076)	(0.072)
Democrat	−0.258**	−0.189**
	(0.060)	(0.044)
Republican	0.040	0.127**
	(0.067)	(0.048)
Independent	−0.074	−0.017
	(0.062)	(0.046)
Constant	−0.093	−0.283**
	(0.075)	(0.081)
N	861	1,903
Adjusted R^2	0.192	0.238

Note: OLS regressions with HC2 robust standard errors in parentheses; YouGov survey data with weights applied.
$^{*}p < 0.05$; $^{**}p < 0.01$.

politics websites—the preponderance of the content encountered is ideologically moderate. There is also suggestive evidence of an intense subgroup of Republicans who, possibly in addition to mainstream sources, consume large quantities of conservative, but not liberal, news and information about politics. Similar bumps on the left correspond to the popular viral site BuzzFeed and other left-leaning mainstream sources, in addition to partisan destinations such as Daily Kos.[14] Arguably, then, most people are not habitual partisan news consumers.

[14]Interestingly, these results diverge somewhat from the analysis of Nelson and Webster (2017), who report a consistent pattern of ideological diversity across news domains of differing levels of partisanship. There are numerous differences between the data sets in this study and the one used by the authors, such as the time period, the number of sources covered, and the measure of site ideology employed. Arguably the two most consequential differences

FIGURE 2 Slant of Site Visits by Respondent Partisanship



Note: Site ideological slant on the x-axis is measured as the average self-reported ideological placement of Facebook users who share "hard news" articles from a given domain. Only visits from pages classified as related to news/politics are included.

But it might *seem* so from the point of view of news publishers, which may lack the ability to see the individuals lurking behind inbound traffic[15]—leading to the possibility of feedback loops via engagement metrics and optimization.

More importantly from the perspective of democratic responsiveness, the minority of individuals with the most polarized political media diets may also be more likely to seek to influence politics by participating in the electoral process, setting the agenda of public discourse, or lobbying their representatives. In this sense, a seemingly small subgroup of the population may serve as opinion leaders who indirectly influence both political leaders and other members of the public. Although such a process has not always been described by scholars as normatively troubling (e.g., Lazarsfeld, Berelson, and Gaudet 1944), the existence of partisan selective exposure among opinion leaders raises the possibility of systematic distortions in elite influence.

To investigate this possibility, I link indicators of verified voter turnout from TargetSmart to the 2016 data, allowing me to identify respondents who voted in the 2016 general and presidential primary elections. I then estimate a single quantity, $E[P(\text{Voted}_i|D = d)]$, across deciles d of average media diet slant in the sample. Table 3 shows that, in fact, turnout was markedly higher among those with the most homogeneously partisan media consumption habits—especially so for those in the ninth and tenth deciles (i.e., the most conservative media diets), where average 2016 general-election turnout is estimated to be 64% and 68%, respectively (compared to an average of 57% for the entire sample) and presidential-primary turnout is 17% and 26% (compared to 14% for the entire sample).[16] In a real sense, then, even if most Americans do not exist in online echo chambers, they are subject to the political influence of those who do.

are the granularity of my data, which contain full URLs to specific news articles, and the number of sources for which I have a measure of media slant. Ultimately, the comScore data employed by Nelson and Webster (2017) are best suited for questions about the audiences of specific domains. Site analytics data are designed to provide representative data on site traffic, whereas the data and analysis in this section aggregate across domains to draw generalizations about the total number of visits to news articles from different partisan perspectives.

[15]A similar dynamic may also explain the enduring appeal of the "echo chambers" narrative. Evidence adduced in favor of the phenomenon often focuses on the poles of the distribution, effectively selecting on the dependent variable. See http://graphics.wsj.com/blue-feed-red-feed/ for an example.

[16]This analysis disregards well-known confounds such as age. Still, the existence of this relationship illustrates that whether or not media diets are a byproduct of other factors, the electorate in 2016 (and likely other years) was disproportionately composed of people with very partisan media consumption patterns.

TABLE 3 2016 Voter Turnout by Average Media Diet Decile

	$E[P(\text{Voted}_i \mid D = d)]$	
Decile	**General**	**Presidential primary**
1	0.585	0.179
2	0.552	0.170
3	0.473	0.124
4	0.561	0.141
5	0.529	0.119
6	0.490	0.083
7	0.589	0.140
8	0.595	0.116
9	0.642	0.168
10	0.684	0.256

Note: TargetSmart verified voter turnout data. Fractions represent weighted means, using YouGov's sample-matching weights, across deciles of respondents' average media diet slant. Individual media diets are computed without using the news/politics classifier in order to maximize the number of observations.

Discussion

Using novel approaches in data and measurement, I find that most news consumers are not partisan in the sources they use to learn about politics. Indeed, browsing for news about politics is a rare activity: No more than 7%–9% of total web visits in either year were to pages with an associated Bakshy et al. alignment score (indicating "hard news" domains), and my classifier predicts that fewer than a quarter of those correspond to politically relevant news articles. People who do regularly read political news, and who rely on partisan sources to do so—a relatively small share of the population—drive a disproportionate amount of traffic to partisan outlets and also appear to participate more in politics. This pattern is consistent with broader findings about the relative lack of political interest in the population and in studies that show a high degree of responsiveness to the most engaged citizens, who are more knowledgeable and consistent in their attitudes (Abramowitz and Saunders 2008; Delli Carpini and Keeter 1996; Lawrence, Sides, and Farrell 2010). A contribution of this study is to show that, knowledge and consistency aside, these opinion leaders are likely to rely on much more ideologically extreme information sources to shape their attitudes and behaviors than the rest of the public.

That the picture is one of relative overall moderation in political news consumption suggests some role for friction or inertia in people's media habits.[17] Despite a wealth of available information from diverse perspectives, many are satisfied with what is in front of them. In this sense, perhaps not so much has changed. Writing from an earlier era, Doris Graber observed that her subjects did not seem particularly selective in their choices of what to watch or listen to on the news: "The casualness of the news selection and rejection process is the main barrier to systematic selectivity. … [O]ur panelists selected information primarily on an opportunity basis. They read the newspapers and magazines that were within convenient physical reach, looking for whatever pleasing stories might be presented on a certain day" (Graber 1988, p. 131).

These findings arose from a tradition in communication research that combined intensive interviews and diaries of people's news habits. Motivating her study of news processing, Graber noted the problems with existing survey-based methods, including socially desirable responding, the limitations of closed-ended questions, and an inability to observe people unobtrusively in naturalistic settings (Graber 1988, pp. 17–21). Opting for in-person, in-depth interviews was a creative solution to these problems, but one that came with a cost: The original study had a sample of 21 subjects. Taking a different approach but with similar goals, this study benefits from passive metering technologies, text analysis, and machine learning to scale up ecologically valid observation of individuals' online political media consumption behavior. Paired with linked survey responses, these data provide important descriptive evidence on the prevalence and sources of online partisan selective exposure. Future work should build on these advances by tracking survey respondents' attitudes and news recall in successive panel waves and supplementing them with content analysis, allowing for a more in-depth understanding of responses to real-world events and how they are covered in the media.

This approach is merely a starting point for a larger research agenda on selective exposure and media effects in the 21st-century, high-choice media ecosystem. Many questions about the consequences and determinants of media choice remain insufficiently understood. Most urgently, what is the role of social media in driving people's information consumption (Barberá 2014; Eady et al. 2019)? Do the pathways by which people actively seek out information affect the type of content they encounter (e.g., Flaxman, Goel, and Rao 2016)? This exploration

[17] A potentially related factor is that other attributes, such as the topic of a news item, may overwhelm any underlying preference for congenial content (Mummolo 2016).

should not be limited to strictly political content or "hard news"; the role of preferences over lifestyle choices and how they potentially contribute to "de facto" selectivity has too rarely been studied (DellaPosta, Shi, and Macy 2015). Finally, a better understanding of the information people expose themselves to will enable well-identified research on the real-world effects of online media content on attitudes, knowledge, and political behavior (Arceneaux and Johnson 2013; de Benedictis-Kessner et al. 2019; Levendusky 2013).

Some of these questions, such as those about the role of social media and mobile app use, will require further innovations in data collection and measurement (e.g., Reeves et al. 2019). However, passive metering data on web visits will continue to be a rich resource for studying exposure to information (and misinformation) and its effects (e.g., Guess, Nyhan, and Reifler 2020). As the quantity and breadth of these data increase, it will be possible to conduct more finely grained analyses of media consumption behavior within particular demographic and political subgroups. More data will also facilitate exploration of over-time dynamics: How predictable and how stable are people's information-seeking behavior (Scharkow et al. 2020)? How do people respond to political events or to changes in the political environment? Can we trace how the introduction of new outlets shifts people's habits? Beyond the realm of political communication, such data may also prove useful for difficult-to-measure constructs such as political interest and digital literacy.

Although this study does not test the effect of media consumption on knowledge about public affairs, it provides evidence directly relevant to concerns about average citizens' exposure to news and information about politics—a commonly held precondition for meaningful democratic deliberation. Contrary to dire predictions that Americans would use the internet to cocoon themselves into mutually exclusive echo chambers, leading to polarization of attitudes and irreconcilable factual beliefs, I instead find a pattern of overlap in people's media consumption patterns. This is due to the relatively balanced nature of most people's media diets for political news, which partially reflects a reliance on large, mainstream online intermediaries (as shown in Appendix F). From the perspective of democratic theory, these websites provide two vital functions: shared mediated experiences and access to diverse perspectives. Although it is still true that people have more opportunities than ever to select out of politics (Prior 2007), this study suggests that in practice, the sometimes-rewarding, sometimes-discomfiting experience of political competition may be difficult to avoid altogether.

References

Abramowitz, A.I., and K.L. Saunders. 2008. "Is Polarization a Myth?" *Journal of Politics* 70(2): 542–55.

Arceneaux, Kevin, and Martin Johnson. 2013. *Changing Minds or Changing Channels? Partisan News in an Age of Choice.* Chicago, IL: University of Chicago Press.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348(6239): 1130–32.

Barberá, Pablo. 2014. "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S." Working paper, last accessed on June 28, 2020 at http://pablobarbera.com/static/barbera_polarization_APSA.pdf.

Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom.* New Haven, CT: Yale University Press.

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign.* Chicago, IL: University of Chicago Press.

Bowyer, Benjamin, and Joseph Kahne. 2019. "Motivated Circulation: How Misinformation and Ideological Alignment Influence the Circulation of Political Content. " *International Journal of Communication* 13:25.

Budak, Ceren, Sharad Goel, and Justin M. Rao. 2016. "Fair and Balanced? Quantifying Media Bias Through Crowdsourced Content Analysis." *Public Opinion Quarterly* 80(S1): 250–71.

Clemons, Traci E., and Edwin L. Bradley Jr. 2000. "A Nonparametric Measure of the Overlapping Coefficient." *Computational Statistics & Data Analysis* 34(1): 51–61.

de Benedictis-Kessner, Justin, Matthew A. Baum, Adam J. Berinsky, and Teppei Yamamoto. 2019. "Persuading the Enemy: Estimating the Persuasive Effects of Partisan Media with the Preference-Incorporating Choice and Assignment Design." *American Political Science Review* 113(4): 902–16.

DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" *American Journal of Sociology* 120(5): 1473–1511.

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

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Dvir-Gvirsman, Shira, Yariv Tzfati, and Ericka Menchen-Trevino. 2014. "The Extent and Nature of Ideological Selective Exposure Online: Combining Survey Responses with Actual Web Log Data from the 2013 Israeli Elections." *New Media & Society* 18(5): 857–77.

Eady, Gregory, Jonathan Nagler, Andy Guess, Jan Zilinsky, and Joshua A. Tucker. 2019. "How Many People Live in Political Bubbles on Social Media? Evidence from Linked Survey and Twitter Data." *Sage Open* 9(1): 2158244019832705.

Fishkin, James S. 1991. *Democracy and Deliberation: New Directions for Democratic Reform.* Cambridge: Cambridge University Press.

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.

Freedman, Jonathan L., and David O. Sears. 1965. “Selective Exposure.” *Advances in Experimental Social Psychology* 2: 57–97.

Garrett, Kelly R. 2009. “Echo Chambers Online?: Politically Motivated Selective Exposure Among Internet News Users.” *Journal of Computer-Mediated Communication* 14(2): 265–85.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. “What Drives Media Slant? Evidence from U.S. 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.

Graber, Doris A. 1988. *Processing the News: How People Tame the Information Tide.* New York: Longman.

Groeling, Tim. 2013. “Media Bias by the Numbers: Challenges and Opportunities in the Empirical Study of Partisan News.” *Annual Review of Political Science* 16: 129–51.

Groseclose, Tim, and Jeffrey Milyo. 2005. “A Measure of Media Bias.” *The Quarterly Journal of Economics* 120(4): 1191–1237.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker. 2019. “Less than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook.” *Science Advances* 5(1): eaau4586.

Guess, Andrew M. 2015. “Measure for Measure: An Experimental Test of Online Political Media Exposure.” *Political Analysis* 23(1): 59–75.

Guess, Andrew M., Brendan Nyhan, and Jason Reifler. 2020. “Exposure to Untrustworthy Websites in the 2016 U.S. Election. ” *Nature Human Behaviour* 4: 472–80.

Habermas, Jürgen. 1984. *The Theory of Communicative Action, Volume I.* Boston: Beacon.

Hindman, Matthew. 2008. *The Myth of Digital Democracy.* Princeton, NJ: Princeton University Press.

Hall Jamieson, Kathleen, and Joseph N. Cappella. 2008. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment.* Oxford: Oxford University Press.

Inman, Henry F., and Edwin L. Bradley Jr. 1989. “The Overlapping Coefficient as a Measure of Agreement Between Probability Distributions and Point Estimation of the Overlap of Two Normal Densities.” *Communications in Statistics – Theory and Methods* 18(10): 3851–74.

Iyengar, Shanto, and Kyu S. Hahn. 2009. “Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use.” *Journal of Communication* 59(1): 19–39.

Iyengar, Shanto, Kyu S. Hahn, Jon A. Krosnick, and John Walker. 2008. “Selective Exposure to Campaign Communication: The Role of Anticipated Agreement and Issue Public Membership.” *The Journal of Politics* 70(1): 186–200.

Katz, Elihu. 1996. “And Deliver Us From Segmentation.” *The Annals of the American Academy of Political and Social Science* 546(1): 22–33.

Klapper, Joseph T. 1960. *The Effects of Mass Communication.* Glencoe, IL: Free Press.

Klasa, Adrienne. 2016. “Breaking the Liberal Echo Chamber.” HuffPost. Last accessed on July 1, 2020 at https://www.huffpost.com/entry/breaking-the-liberal-echo-chamber_b_582492a9e4b0270d9a2ad833.

Lawrence, E., J. Sides, and H. Farrell. 2010. “Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics.” *Perspectives on Politics* 8(1): 141–57.

Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. 1944. *The People’s Choice: How the Voter Makes Up His Mind in a Presidential Campaign.* New York: Columbia University Press.

Levendusky, Matthew S. 2013. “Why Do Partisan Media Polarize Viewers?” *American Journal of Political Science* 57(3): 611–23.

Lelkes, Yphtach. 2016. “Mass Polarization: Manifestations and Measurements.” *Public Opinion Quarterly* 80(S1): 392–410.

Lorenz-Spreen, Philipp, Stephan Lewandowsky, Cass R. Sunstein, and Ralph Hertwig. 2020. “How Behavioural Sciences Can Promote Truth, Autonomy and Democratic Discourse Online.” *Nature Human Behaviour* 4(11): 1102–09.

Madrigal, Alexis. 2017. “What Facebook Did to American Democracy.” *The Atlantic*. Last accessed on July 1, 2020 at https://www.theatlantic.com/technology/archive/2017/10/what-facebook-did/542502/.

Messing, Solomon, and Sean J. Westwood. 2014. “Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online.” *Communication Research* 41(8): 1042–63.

Mutz, Diana C. 2006. *Hearing the Other Side: Deliberative versus Participatory Democracy.* Cambridge: Cambridge University Press.

Mummolo, Jonathan. 2016. “News from the Other Side: How Topic Relevance Limits the Prevalence of Partisan Selective Exposure.” *The Journal of Politics* 78(3): 763–73.

Negroponte, Nicholas. 1995. *Being Digital.* New York: Alfred A. Knopf.

Nelson, Jacob L., and James G. Webster. 2017. “The Myth of Partisan Selective Exposure: A Portrait of the Online Political News Audience.” *Social Media + Society* 3(3): 2056305117729314.

Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You.* UK: Penguin.

Perryman, Mallory R. 2017. Public Perceptions of Partisan Selective Exposure. PhD thesis, The University of Wisconsin–Madison.

Peterson, Erik, Sharad Goel, and Shanto Iyengar. 2019. “Partisan Selective Exposure in Online News Consumption: Evidence from the 2016 Presidential Campaign.” *Political Science Research and Methods* 1–17.

Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections.* Cambridge: Cambridge University Press.

Prior, Markus. 2013. “Media and Political Polarization.” *Annual Review of Political Science* 16: 101–27.

Reeves, Byron, Nilam Ram, Thomas N. Robinson, James J. Cummings, C. Lee Giles, Jennifer Pan, Agnese Chiatti, M.J. Cho, Katie Roehrick, Xiao Yang et al. 2019. “Screenomics: A Framework to Capture and Analyze Personal Life

Experiences and the Ways that Technology Shapes Them." *Human–Computer Interaction* 36(2): 150–201.

Scharkow, Michael, Frank Mangold, Sebastian Stier, and Johannes Breuer. 2020. "How Social Network Sites and Other Online Intermediaries Increase Exposure to News." *Proceedings of the National Academy of Sciences* 117(6): 2761–63.

Shapiro, Robert Y. 2013. "Hearing the Opposition: It Starts at the Top." *Critical Review* 25(2): 226–44.

Shmargad, Yotam, and Samara Klar. 2020. "Sorting the News: How Ranking by Popularity Polarizes Our Politics." *Political Communication* 37(3): 423–46.

Stier, Sebastian, Johannes Breuer, Pascal Siegers, and Kjerstin Thorson. 2019. "Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field." *Social Science Computer Review* 38(5): 503–16.

Stroud, N.J. 2008. "Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure." *Political Behavior* 30(3): 341–66.

Stroud, N.J. 2011. *Niche News: The Politics of News Choice.* Oxford: Oxford University Press.

Sunstein, Cass R. 2017. *#Republic.* Princeton, NJ: Princeton University Press.

Sunstein, Cass R., and Richard Thaler. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* New Haven, CT: Yale University Press.

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.

Valentino, Nicholas A., Antoine J. Banks, Vincent L. Hutchings, and Anne K. Davis. 2009. "Selective Exposure in the Internet Age: The Interaction between Anxiety and Information Utility." *Political Psychology* 30(4): 591–613.

Van Alstyne, Marshall, and Erik Brynjolfsson. 2005. "Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities." *Management Science* 51(6): 851–68.

Walker, Mason. 2019. "Americans favor mobile devices over desktops and laptops for getting news." Pew Research Center report, November 19, 2019. Downloaded June 9, 2020 from https://www.pewresearch.org/fact-tank/2019/11/19/americans-favor-mobile-devices-over-desktops-and-laptops-for-getting-news/.

Weeks, Brian E., Daniel S. Lane, Dam Hee Kim, Slgi S. Lee, and Nojin Kwak. 2017. "Incidental Exposure, Selective Exposure, and Political Information Sharing: Integrating Online Exposure Patterns and Expression on Social Media." *Journal of Computer-Mediated Communication* 22(6): 363–79.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix A: Additional Tables and Figures
Appendix B: Details and Validation of Survey and Web Data
Appendix C: Estimating Web Domain Slant
Appendix D: Details on Political Article Classifier
Appendix E: Additional Regression Models
Appendix F: Portals as Information Intermediaries
Appendix G: Mainstream Media Sites and News Consumption