NBER WORKING PAPER SERIES

IDEOLOGICAL SEGREGATION ONLINE AND OFFLINE

Matthew Gentzkow
Jesse M. Shapiro

Working Paper 15916
http://www.nber.org/papers/w15916

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2010

This paper would not have been possible without the generous support of Jim Collins at Mediamark Research and Intelligence. We thank our dedicated research assistants for invaluable contributions to this project, and seminar participants at Chicago Booth and the SIEPR / Microsoft Conference on Internet Economics for helpful comments. This research was funded by the Initiative on Global Markets, the George J. Stigler Center for the Study of the Economy and the State, the Centel Foundation / Robert P. Reuss Faculty Research Fund, and the Neubauer Family Foundation, all at the University of Chicago Booth School of Business. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Ideological Segregation Online and Offline
Matthew Gentzkow and Jesse M. Shapiro
NBER Working Paper No. 15916
April 2010
JEL No. D83,L86

ABSTRACT

We use individual and aggregate data to ask how the Internet is changing the ideological segregation of the American electorate. Focusing on online news consumption, offline news consumption, and face-to-face social interactions, we define ideological segregation in each domain using standard indices from the literature on racial segregation. We find that ideological segregation of online news consumption is low in absolute terms, higher than the segregation of most offline news consumption, and significantly lower than the segregation of face-to-face interactions with neighbors, co-workers, or family members. We find no evidence that the Internet is becoming more segregated over time.

Matthew Gentzkow
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
gentzkow@chicagobooth.edu

Jesse M. Shapiro
University of Chicago
Booth School of Business
5807 S. Woodlawn Avenue
Chicago, IL 60637
and NBER
jmshapir@uchicago.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w15916

1 Introduction

Democracy is most effective when citizens have accurate beliefs. To form such beliefs, individuals must encounter information which will sometimes contradict their pre-existing views. Guaranteeing exposure to information from diverse viewpoints has been a central goal of media policy in the United States and around the world (Gentzkow and Shapiro 2008).

New technologies such as the Internet could either increase or decrease the likelihood that consumers are exposed to diverse news and opinion. The Internet dramatically reduces the cost of acquiring information from a wide range of sources. But increasing the number of available sources can also make it easier for consumers to self-segregate ideologically, limiting themselves to those that are likely to confirm their prior views (Mullainathan and Shleifer 2005).

The possibility that the Internet may be increasing ideological segregation has been articulated forcefully by Sunstein (2001): "Our communications market is rapidly moving" toward a situation where "people restrict themselves to their own points of view—liberals watching and reading mostly or only liberals; moderates, moderates; conservatives, conservatives; Neo-Nazis, Neo-Nazis" (4-5). This limits the "unplanned, unanticipated encounters [that are] central to democracy itself" (9). Sunstein (2001) also notes that the rise of the Internet will be especially dangerous if it crowds out other activities where consumers are more likely to encounter diverse viewpoints. He argues that both traditional media such as newspapers, magazines, and broadcasters, and face-to-face interactions in workplaces and local communities are likely to involve such diverse encounters.[1]

In this paper, we assess the extent to which news consumption on the Internet is ideologically segregated, and compare online segregation to segregation of both traditional media and face-to-face interactions. For each outlet in our sample (a newspaper, a particular website), we measure the share conservative: the share of users who report their political outlook as "conservative," among those who report being either "conservative" or "liberal." We then define each individual's *conservative exposure* to be the average share conservative on the outlets she visits. For example, if

[1]"People who rely on [newspapers, magazines, and broadcasters] have a range of chance encounters... with diverse others, and also exposure to materials and topics that they did not seek out in advance" (Sunstein 2001, 11). "The diverse people who walk the streets and use the parks are likely to hear speakers' arguments about taxes or the police; they might also learn about the nature and intensity of views held by their fellow citizens.... When you go to work or visit a park... it is possible that you will have a range of unexpected encounters" (30).

the only outlet an individual visits is nytimes.com, her exposure is defined as the share conservative on nytimes.com. If she visits both nytimes.com and foxnews.com, her exposure is the average of the conservative shares on these two sites. Our main measure of segregation is the "isolation index" (White 1986, Cutler et al. 1999), a standard metric in the literature on racial segregation. In our context, the isolation index is equal to the average conservative exposure of conservatives minus the average conservative exposure of liberals. If conservatives only visit foxnews.com and liberals only visit nytimes.com, the isolation index will be equal to 100 percentage points. If both conservatives and liberals get all their news from cnn.com, the two groups will have the same conservative exposure, and the isolation index will be equal to zero.

We use aggregate 2009 data on website audiences from comScore, supplemented with micro data on the browsing behavior of individuals from 2004-2008. To measure offline consumption, we use 2008 individual-level data from Mediamark Research and Intelligence on consumption of newspapers, magazines, broadcast television, and cable. To measure face-to-face interactions, we use data on the political views of individuals' friends and acquaintances as reported in the 2006 General Social Survey.

News consumption online is far from perfectly segregated. The average Internet news consumer's exposure to conservatives is 57 percent, slightly to the left of the US adult population. The average conservative's exposure is 60.6 percent, similar to a person who gets all her news from usatoday.com. The average liberal's exposure is 53.1 percent, similar to a person who gets all her news from cnn.com. The isolation index for the Internet is 7.5 percentage points, the difference between the average conservative's exposure and the average liberal's exposure.

News consumers with extremely high or low exposure are rare. A consumer who got news exclusively from nytimes.com would have a more liberal news diet than 95 percent of Internet news users, and a consumer who got news exclusively from foxnews.com would have a more conservative news diet than 99 percent of Internet news users.

The isolation index we estimate for the Internet is higher than that of broadcast television (1.8), magazines (2.9), cable television (3.3), and local newspapers (4.1), and lower than that of national newspapers (10.4). We estimate that eliminating the Internet would reduce the ideological segregation of news and opinion consumption across all media from 4.9 to 3.8.

Online segregation is somewhat higher than that of a social network where individuals matched randomly within counties (5.9), and lower than that of a network where individuals matched randomly within zipcodes (9.4). It is significantly lower than the segregation of actual networks formed through voluntary associations (14.5), work (16.8), neighborhoods (18.7), or family (24.3). The Internet is also far less segregated than networks of trusted friends (30.3).

Using our micro data sample, we estimate online segregation back to 2004, and find no evidence that the Internet is becoming more segregated over time.

We explore two economic mechanisms that limit the extent of online segregation. First, online news is vertically differentiated, with most consumption concentrated in a small number of relatively centrist sites. Much of the previous discussion of Internet segregation has focused on the "long tail" of political blogs, news aggregators, and activist sites. We confirm that these sites are often ideologically extreme, but find that they account for a very small share of online consumption. Second, a significant share of consumers get news from multiple outlets. This is especially true for visitors to small sites such as blogs and aggregators. Visitors of extreme conservative sites such as rushlimbaugh.com and glennbeck.com are more likely than a typical online news reader to have visited nytimes.com. Visitors of extreme liberal sites such as thinkprogress.org and moveon.org are more likely than a typical online news reader to have visited foxnews.com.

In the final section of results, we ask how segregation at the level of individual stories may differ from segregation at the level of the news outlet. The two could differ if liberals and conservatives choose different content within a given outlet. In daily newspapers, for example, conservatives and liberals might both read the *Wall Street Journal*, but conservatives might concentrate on the editorial pages while liberals concentrate on the news section. To gauge the importance of this kind of sorting on the Internet, we present evidence from case studies of two major news events– the Virginia Tech shootings in 2007 and the presidential election in 2008. On both of these days, the number of hits to news websites spikes significantly, and most content consumed presumably focuses on these major events. The isolation index for these days, however, is if anything lower than on an average day. These cases provide some evidence that online segregation is low even when within-outlet sorting is limited, and that conservatives and liberals are not highly segregated in their sources for information about major news events.

We conclude with an important caveat: none of the evidence here speaks to the way people translate the content they encounter into beliefs. People with different ideologies see similar content, but both Bayesian (Gentzkow and Shapiro 2006; Acemoglu et al. 2009) and non-Bayesian (Lord et al. 1979) mechanisms may lead people with divergent political views to interpret the same information differently.

Our results inform both popular and theoretical discussions of the political impact of the increased media competition. Mullainathan and Shleifer (2005), Sobbrio (2009), and Stone (2010) write down theoretical models of media markets in which increasing the number of outlets may lead consumers to become more segregated ideologically. Public officials (e.g., Leibowitz 2010) and commentators (e.g., Brooks 2010) routinely warn of the dangerous effects of ideological isolation in news consumption on the health of our democracy. Sunstein (2001), Kohut (2004), Von Drehle (2004), Carr (2008), and Friedman (2009), among others, have argued that proliferation of news sources on the Internet may be increasing that isolation.

To our knowledge, ours is the first study to use detailed data on the ideological composition of news-website visitors to compare ideological segregation online and offline. The best existing evidence on ideological segregation online uses data on patterns of links rather than consumption (Adamic and Glance 2005). Tewksbury (2005) presents evidence on demographic (not specifically ideological) specialization in online news audiences.

A large literature considers the causes and effects of political polarization (McCarty et al 2006; Glaeser and Ward 2006), which Campante and Hojman (2010) relate to the structure of the media market. A growing literature in economics studies the effects of the news media on public policy (e.g., Stromberg 2004, Stromberg and Snyder forthcoming), political beliefs and behavior (Prior 2005, Gentzkow 2006, DellaVigna and Kaplan 2007, Knight and Chiang 2008), and social capital (Olken 2009).

Section 2 below describes the data used in our study. Section 3 introduces our segregation measure and empirical strategy. Section 4 presents our main results. Section 5 discusses economic explanations of our findings and section 6 discusses segregation of content (as opposed to site) viewership. Section 7 presents robustness checks. Section 8 concludes.

2 Data

2.1 Internet News

Our Internet news data are provided by comScore.

To construct our universe of national political news and opinion websites, we begin with all sites that comScore categorizes as "General News" or "Politics." We exclude sites of local newspapers and television stations, other local news and opinion sites, and sites devoted entirely to non-political topics such as sports or entertainment. We supplement this list with the sites of the 10 largest US newspapers (as defined by the Audit Bureau of Circulations for the first half of 2009). We also add all domains that appear on any of thirteen online lists of political news and opinion websites.[2] The final list includes 1,379 sites.

We measure site size using the average daily unique visitors to each site over the twelve months in 2009 from comScore MediaMetrix. MediaMetrix data come from comScore's panel of over one million US-resident Internet users. Panelists install software on their computers to permit monitoring of their browsing behavior, and comScore uses a passive method to distinguish multiple users of the same machine. Media Metrix only reports data for sites that were visited by at least 30 panelists in a given month. We have at least one month of Media Metrix data for 459 of the sites on our list.

We measure site ideological composition as the share of daily unique visitors who are conservative over the twelve months in 2009 from comScore PlanMetrix. PlanMetrix data come from a survey distributed electronically to approximately 12,000 comScore panelists. The survey asks panelists the question "In terms of your political outlook, do you think of yourself as...? [very conservative / somewhat conservative / middle of the road / somewhat liberal / very liberal]". We classify those who answer "middle of the road" as missing data and we classify all others as either

[2] These lists are rightwingnews.com's "100 Of The Most Popular Political Websites On The Net", "The Blogosphere Power Rankings – The Most Popular Political Blogs On The Net", and "The Top 125 Political Websites On The Net Version 5.0"; alexa.com's "Top Sites News > Weblogs" and "Politics News"; evancarmichael.com's "Top 50 Political Blogs: 2009"; intellectualconservative.com's "Top 100 Conservative Political Websites of 2007" and "Top 100 Liberal Political Websites of 2007"; wikio.com's "Top Blogs - Politics"; urbanconservative.com's "The Best Conservative Blogs on the Internet – Period!"; reachm.com/amstreet's "Top 100 Liberal Bloggers or Sites, by traffic as of 12/19/07"; politicalbloglistings.blogspot.com's "List of Political Blogs"; and toppoliticalsites.org's "Top Political Sites". We exclude any sites for which the lists provide several URLs for one domain name, where the URL is a subdomain (e.g., newscompass.blogspot.com), or where the top level domain does not provide news or opinion content (e.g., twitter.com).

conservative or liberal. Section 4.2 presents detailed results on exposure for all five categories, and section 7.3 reports isolation measures treating "middle of the road" panelists as conservative or liberal.

PlanMetrix data are only available for relatively large sites. We have at least one month of Plan Metrix data on ideological composition for 119 of the sites on our list. This set of sites forms our primary sample.

To perform robustness checks and to measure changes over time, we use comScore microdata on the browsing behavior of a subset of panelists obtained from Wharton Research Data Services (WRDS). We have separate data extracts for 2004, 2006, 2007, and 2008. The data include 50,000-100,000 machines and contain the domain name of each site visited.

The data include the zipcode where each machine is located. From this, we construct a proxy for ideology, which is a dummy for whether the share of political contributions going to Republicans from 2000-2008 in the zipcode is above the national median. We construct this variable from Federal Election Commission data on political contributions as in Gentzkow and Shapiro (2010).

Relative to the site-level aggregates, the microdata have two important limitations. First, because the comScore microdata are defined at the domain level (e.g., yahoo.com), we cannot distinguish news content on sub-pages of large sites such as aol.com and yahoo.com. Sites such as Yahoo! News and AOL News are therefore excluded from the microdata sample. Second, the microdata do not distinguish between multiple users of the same machine.

2.2 Offline Media

Our data on offline media consumption are provided by Mediamark Research & Intelligence (MRI).

We use data on 51,354 respondents from the spring 2007 and spring 2008 waves of the MRI Survey of the American Consumer.

Data on cable television comes from questions asking the number of hours respondents viewed CNN, Fox News, MSNBC, CNBC, and Bloomberg cable networks respectively in the last 7 days. We estimate the number of days each respondent viewed each network in the last 7 days by assuming one hour of viewing per viewing day and top-coding at 7 days of viewing where necessary.

Data on broadcast television comes from questions asking the number of days in the last 5 weekdays respondents viewed the evening newscasts of ABC, CBS, NBC, PBS or the BBC (which is broadcast in some markets on public television stations) respectively.

Data on national newspapers come from questions asking whether respondents read the most recent weekday edition of *The New York Times*, *USA Today*, and *The Wall Street Journal* respectively.

Data on magazines come from questions asking whether respondents read the most recent issue of *The Atlantic*, *Barron's*, *BusinessWeek*, *The Economist*, *Forbes*, *Fortune*, *The New Yorker*, *Newsweek*, *Time*, and *U.S. News & World Report* respectively. We assume that each issue of a magazine is read on one day to convert this to a measure of daily readership.

Data on local newspapers come from a free response question asking which newspapers the respondent read in the last 24 hours. The data extract aggregates this variable into an indicator for whether the respondent read one of the 100 largest U.S. dailies in the last 24 hours. We code a respondent as reading a local newspaper if she read one of the top 100 papers in the last 24 hours but did not report reading one of the national papers in the same window of time. We define a newspaper market as either a PMSA or a county (for counties that are not in PMSAs) and assume that respondents in the same newspaper market who read a local paper read the same paper. Gentzkow and Shapiro (2010) present evidence in support of this market definition.

The MRI survey includes the question "In terms of your political outlook, do you think of yourself as...? [very conservative / somewhat conservative / middle of the road / somewhat liberal / very liberal]," which we use to define each respondent's political ideology as conservative or liberal, as in the comScore data.

The MRI data extract identifies the respondent's zipcode. We use this information to study geographic segregation in ideology, as a supplement to the data on face-to-face interactions described in section 2.3 below.

The MRI data extract includes sampling weights to account for their multistage sample selection process. We use these weights in our main analysis and present unweighted results as a robustness check in the online appendix. MRI also imputes missing values for a section of the survey that includes the political ideology question; we treat these respondents as having missing ideology data.

2.3 Face-to-Face Interactions

Our data on face-to-face interactions come from the 2006 wave of the General Social Survey (GSS).

The 2006 wave of the GSS includes a "Number Known" topical module, which DiPrete et al. (forthcoming) designed to measure segregation in social networks. A total of 1,347 respondents answered one or more questions in this module.

Respondents are asked about the characteristics (race, religiosity, etc.) of their family members, friends, and acquaintances. For each group, respondents are asked the number they are "pretty certain are strongly liberal" and "pretty certain are strongly conservative." Responses are categorical: 0, 1, 2-5, 6-10, more than 10. We recode these responses at the midpoint of the respective category with an arbitrary topcode of "12" for the largest category. In the online appendix we present results excluding respondents with topcoded responses.

We use data for the following groups: (i) the respondent's family; (ii) the respondent's neighborhood; (iii) the respondent's workplace; (iv) people the respondent is acquainted with via clubs, schools, associations, or places of worship; (v) people the respondent trusts. We define the share conservative for each group to be the number the respondent identifies as strongly conservative divided by the share identified as either strongly conservative or strongly liberal.

Data on respondents' political ideology comes from the question "I'm going to show you a seven-point scale on which the political views that people might hold are arranged from extremely liberal–point 1–to extremely conservative–point 7. Where would you place yourself on this scale?" We classify respondents as either conservative or liberal, treating moderates (point 4) as having missing ideology data.

We weight data using the GSS's WTSS weight variable which accounts for re-sampling of non-respondents and the presence of multiple adults per household. In the online appendix, we present results weighting respondents equally.

2.4 Comparability of Online and Offline Sources

Both comScore and MRI are highly regarded proprietary sources for information on the size and composition of media audiences.

To confirm the comparability and validity of the two sources, we exploit the fact that the MRI survey asks respondents whether they got news online from ABC News, AOL News, CBS News, CNN, Fox News, MSNBC, the New York Times, USA Today, the Wall Street Journal, or Yahoo! News in the last 24 hours.

Figure 1 shows that political outlook in the MRI and comScore data match closely. The number of daily visits is also highly correlated between the two sources ($\rho > 0.9$).

3 Measuring Ideological Segregation

3.1 Definition

Let $m \in M$ index "media" (Internet, broadcast news, etc., as well as domains of face-to-face interaction such as zipcodes or workplaces). Let $j \in J$ index individual "outlets" (cnn.com, ABC Nightly News, etc., or a particular zipcode, workplace, etc.). The set J is partitioned into mutually exclusive subsets J_m, the set of outlets j in medium m.

Let $i \in I$ index individuals. Each individual is either liberal or conservative.

Let $v_{ij} \in \{0,1\}$ indicate whether a given individual i visits outlet j. For news media such as the Internet, a given individual may visit multiple outlets. For domains of face-to-face interaction such as zipcodes, each individual "visits" one and only one outlet.

Define $cons_j$ and lib_j to be the number of conservative and liberal visits respectively to outlet j. Define $cons_m$ and lib_m to be the total number of conservative and liberal visits on medium m, and define $visits_j = cons_j + lib_j$ and $visits_m = cons_m + lib_m$.

Our primary measure of segregation is the isolation index (White 1986, Cutler et al. 1999). For a given medium m this is:

$$S_m \equiv \frac{\sum_{j \in J_m} \left(\frac{cons_j}{cons_m} \cdot \frac{cons_j}{visits_j} \right) - \left(\frac{cons_m}{visits_m} \right)}{1 - \left(\frac{cons_m}{visits_m} \right)}. \tag{1}$$

S_m is symmetric in the sense that it would be unchanged if we replaced $cons_j$ and $cons_m$ in the definition with lib_j and lib_m.

We refer to $\frac{cons_j}{visits_j}$ as the *share conservative* of site j, and we refer to the average share conservative on outlets that i visits as i's *conservative exposure.* The first term in the numerator, $\sum_{j\in J_m}\left(\frac{cons_j}{cons_m}\cdot\frac{cons_j}{visits_j}\right)$, is then the visit-weighted average exposure of conservatives.

The isolation index captures the extent to which conservatives disproportionately visit outlets whose other visitors are conservative. The numerator subtracts from the average conservative exposure of conservatives the average conservative exposure if medium m consisted of a single outlet j. The denominator scales the numerator so that S_m ranges from zero (all conservative and liberal visits are to the same outlet) to one (conservatives only visit 100% conservative outlets and liberals only visit 100% liberal outlets).

There are three equivalent ways to write the isolation index. First, S_m is equal to the average conservative exposure of conservatives minus the average conservative exposure of liberals:

$$S_m = \sum_{j\in J_m}\left(\frac{cons_j}{cons_m}\cdot\frac{cons_j}{visits_j}\right) - \sum_{j\in J_m}\left(\frac{lib_j}{lib_m}\cdot\frac{cons_j}{visits_j}\right). \tag{2}$$

Second, S_m is related to the extent to which individuals encounter consumers of the opposite ideology. Suppose that when an individual visits a site she "interacts" with a random visitor to the site. Define D to be the share of all such interactions that are cross-ideology, and define D^{max} to be the maximum possible value of this share–the value it would take if all conservatives and liberals visits were on a single outlet. It is straightforward to show that:

$$S_m = 1 - \frac{D}{D^{max}}. \tag{3}$$

Third, S_m can be written as the share of the variance in ideology that is between sites (as opposed to across individuals within sites). Define I_{jk} to be an indicator for whether the k^{th} visit to outlet j was by a conservative. Then we can write S_m as:

$$S_m = \frac{Var_j\left[E\left(I_{jk}|j\right)\right]}{Var\left(I_{jk}\right)}. \tag{4}$$

In appendix A, we show that the qualitative pattern of our results is the same for two other common segregation measures: the dissimilarity index (Cutler et al. 1999) and the Atkinson index (Frankel and Volij 2008).

3.2 Estimation

Consider a sample of individuals i, with I_{cons} and I_{lib} denoting the samples of individuals with known conservative or liberal ideology, respectively. Let $\hat{m}_j$ be the share of visits to outlet j by consumers with unknown ideology.

For media and geographic units, we compute sample analogues $\widehat{cons}_j$ and $\widehat{lib}_j$ by summing v_{ij} for $i \in I_{cons}$ and $i \in I_{lib}$ respectively, and scaling the resulting values by $1/\left(1-\hat{m}_j\right)$. We compute $\widehat{cons}_m$ and $\widehat{lib}_m$ by summing $\widehat{cons}_j$ and $\widehat{lib}_j$ over j.

For social groupings, we use the individual's report to construct $\widehat{cons}_j$ and $\widehat{lib}_j$ for the relevant group (e.g., acquaintances at work), with each group j corresponding to a particular individual i.[3]

Equation (2) shows that S_m is equal to the average exposure of conservatives minus the average exposure of liberals. We construct an estimate $\hat{S}_m$ equal to the sample analogue of this difference:

$$\hat{S}_m = \frac{1}{\widehat{cons}_m} \sum_{j \in J_m} \sum_{i \in I_{cons}} \left(v_{ij} \cdot \frac{\widehat{cons}_j - v_{ij}}{\widehat{cons}_j + \widehat{lib}_j - v_{ij}} \right) - \frac{1}{\widehat{lib}_m} \sum_{j \in J_m} \sum_{i \in I_{lib}} \left(v_{ij} \cdot \frac{\widehat{cons}_j}{\widehat{cons}_j + \widehat{lib}_j - v_{ij}} \right). \quad (5)$$

Using $\frac{\widehat{cons}_j - v_{ij}}{\widehat{visits}_j - v_{ij}}$ and $\frac{\widehat{cons}_j}{\widehat{visits}_j - v_{ij}}$ as estimates of $\frac{cons_j}{visits_j}$ avoids a finite-sample bias that arises from treating each individual as having been "exposed" to herself.[4]

We estimate the sampling variability in $\hat{S}_m$ using a bootstrap. We report the results in the online appendix and omit them from the main tables because the sampling variability in $\hat{S}_m$ is small enough that it does not affect the comparisons we make.

We will use comScore microdata to study sites and time periods that are outside of our main sample. Let Z_m be identical to S_m, except that $cons_j$ and lib_j are now defined to be the number of consumers from right-of-median and left-of-median zipcodes respectively. To define segregation

[3] The sample of individuals we consider in the GSS is the sample of respondents, rather than the sample of respondents' friends and acquaintances. In the online appendix we report results that treat a respondents' acquaintances as exposed to one another. The latter specification is similar in spirit to DiPrete et al. (forthcoming), who define segregation to be the extent of overdispersion in the "number known" of a given type of person, relative to a benchmark of random network formation. They show that the measure they use is closely related to the isolation index that we use as our primary measure of segregation, though the two measures are reported in different units.

[4] In the appendix A, we show that our main results are strengthened if we ignore this small sample bias. The version of $\hat{S}_m$ that does not make this correction is biased upward in small samples, because the expectation of $\frac{\widehat{cons}_j}{\widehat{cons}_j + \widehat{lib}_j}$ conditional on a conservative in our sample visiting j is greater than its true expectation, and the expectation of $\frac{\widehat{cons}_j}{\widehat{cons}_j + \widehat{lib}_j}$ conditional on a liberal in our sample visiting j is less than its true expectation. Both our estimator and the uncorrected one are consistent as $\widehat{visits}_j$ grows large.

when we include the sites, we need to rescale this ideology measure so that it is in the same units as our main measure. Let $\hat{Z}_m$ be the sample analogue defined by equation (5). We assume that the ratio S_m/Z_m is constant across subsets of sites and time periods, and define

$$\hat{S}_m = \hat{Z}_m \frac{\hat{S}'}{\hat{Z}'}, \tag{6}$$

where $\hat{S}'$ and $\hat{Z}'$ are estimates of S and Z from a sample of sites in which we observe both zipcode ideology and reported political outlook. Because our most recent microdata sample is for 2008, we estimate $\frac{\hat{S}'}{\hat{Z}'}$ using 2009 aggregate data and 2008 microdata.

3.3 Discussion

The simplest way to interpret S_m is as a descriptive measure of the extent to which the news diets of conservatives and liberals are systematically different. A world with "liberals watching and reading mostly or only liberals" (Sunstein 2001, 4-5), and conservatives behaving analogously, would imply S_m close to one. A world where preferences for news are mostly independent of ideology would imply S_m close to zero.

We can also think of S_m as a proxy for the extent to which liberals and conservatives are exposed to different facts and opinions. Gentzkow and Shapiro (2010) find that the ideological slant of U.S. daily newspapers is increasing (and roughly linear) in the ideology of the average reader. If this relationship holds for all media, equation 2 implies that S_m is proportional to the difference between the average slant that conservatives consume and the average slant that liberals consume.

Finally, although direct interaction among consumers on news websites is relatively infrequent, it does occur. Equation 3 shows that there is a tight relationship between S_m and the share of such interactions that we would expect to be cross-ideology.

One important point to highlight is that S_m measures the segregation of visits rather than of individuals. Individuals who make more total visits get more weight in the calculation than those who make few. The distinction is irrelevant for geographic segregation, where each person "visits" one and only one neighborhood. But it can matter for media consumption. Although user-weighted segregation is the concept we would ideally like to measure, we cannot calculate it for the Internet using the aggregate data that constitutes our main source. In section 7.1, we use the comScore microdata to estimate the segregation of Internet users and compare it to the segregation of Internet visits.

A second important point concerns the level of time aggregation in defining a visit. We define an Internet visit to mean visiting a given site at least once on a particular day. One could define alternative segregation measures at higher levels of aggregation (weekly or monthly unique visitors) or lower levels of aggregation (unique visitors in a given hour or minute). The distinction is not trivial, because—under the plausible assumption that a group with a high probability of visiting a site within a given time interval also spends more time on the site conditional on visiting in that interval—measured segregation will be higher the lower the level of aggregation.

We choose daily unique visitors for the Internet because it most closely approximates what we can measure for other media. In section 7.2, we discuss the time aggregation issue further and argue that our conclusions are robust to using coarser or finer measures.

4 Main Results

4.1 Segregation Online and Offline

In table 1, we report the estimated share conservative for US adults and the different media in our sample. Based on the MRI survey, we estimate that 67 percent of all adults who report an ideology say they are conservative. (Note that self-described conservatives outnumber self-described liberals in both the General Social Survey and the National Election Study.) The share conservative on different media are similar to the overall population, with cable attracting a slightly larger share of conservatives, and magazines, national newspapers, and the Internet all attracting relatively more liberals. The table also shows that the Internet remains a relatively small share of overall news consumption.

Table 2 shows the estimated share conservative for selected online outlets in our sample. The top of the table shows the ten largest Internet sites, the ten most conservative sites, and the ten most liberal sites. The largest sites are Yahoo! News, AOL News, and msnbc.com, which all attract fairly representative audiences of Internet users (55 percent, 62 percent, and 57 percent conservative respectively). The most conservative sites are billoreilly.com, rushlimbaugh.com, and glennbeck.com, all personal sites of conservative radio or television hosts. We estimate these sites' visitors to be more than 98 percent conservative. The most liberal sites are thinkprogress.org (a liberal blog, 6 percent conservative), blogcritics.org (a blog and news aggregation site, 17 percent conservative), and bvblackspin.com (a blog hosted on AOL's Black Voices site, 17 percent conservative).

Table 3 shows the pattern of share conservative across offline media. Viewers of Fox News cable network are more conservative than viewers of CNN or MSNBC. Viewership of the major network newscasts is fairly representative of the population, while BBC and PBS newscasts attract more liberal viewers. Readers of the *New Yorker* and the *Atlantic* are relatively liberal, while readers of *Barron's* are relatively conservative. Readers of the *New York Times* print edition are substantially more liberal than those of *USA Today* or the *Wall Street Journal*. Quantitatively, offline audiences may be less polarized than some would have suspected. One fifth of Fox News' audience is liberal, and 33 percent of *New York Times* readers are conservative. Consistent with the view that the Internet will increase segregation, the most extreme Internet sites are far more polarized than any source offline.

We present our main estimates of segregation in table 4. The conservative exposure of conservatives on the Internet is 60.6 percent. The conservative exposure of liberals on the Internet is 53.1 percent. The isolation index for the Internet is therefore 60.6-53.1 = 7.5 percentage points. The data clearly reject the view that liberals only get news from a set of liberal sites and conservatives only get news from a set of conservative sites.

The Internet falls near the top of the distribution of segregation for media. Broadcast news is the least segregated (1.8), followed by magazines (2.9) and cable (3.3), then local newspapers (4.1), the Internet (7.5), and national newspapers (10.4).

Weighting these results by the overall size of the different media shown in table 1, we estimate that the isolation index for all media combined is 4.9. Holding the distribution of offline media consumption constant, we estimate that removing the Internet would reduce this number to 3.8.

Face-to-face interactions tend to be more segregated than news media. Random interactions within a respondent's zipcode are more segregated (9.4) than the Internet, though slightly less so than national newspapers. Interactions with acquaintances formed through voluntary associations

(14.5), workplaces (16.8), neighborhoods (18.7), and families (24.3) are more segregated than any news medium, as are interactions with trusted acquaintances (30.3).

Figure 2 shows the same estimates in a different way. Ideological segregation on the Internet is clearly similar to segregation on other media, and substantially smaller than the segregation of face-to-face interactions.

4.2 Distribution of Online Exposure across Consumers

The isolation index captures the segregation of the average visit. To examine other moments of the distribution, we use the comScore microdata to calculate each individual's conservative exposure: the mean of the estimated share conservative across the sites the individual visits.

Figure 3 plots the distribution of conservative exposure across individuals. Half of individuals have conservative exposure between 51 percent and 61 percent. The 95th percentile of the distribution is 76 percent and the 5th percentile is 40 percent.

For comparison, someone who gets all her news from foxnews.com has a conservative exposure of 88 percent, putting her at the 99th percentile. Someone who gets all her news from nytimes.com has a conservative exposure of 40 percent, putting her at the 5th percentile. The vast majority of consumers, therefore, are far from having an exclusively conservative or exclusively liberal news diet.

Table 5 presents exposure between detailed ideology groups. (Exposure is computed analogously to equation 5.) Very liberal individuals have an exposure of 13 percent to other very liberal individuals and 15 percent to very conservative individuals. Very conservative individuals have an exposure of 9 percent to very liberal individuals and 25 percent to very conservative individuals. Exposure across ideological lines is common even for individuals with strongly-held political ideologies.

4.3 Changes in Online Segregation over Time

Table 6 shows how segregation of the Internet has changed over time. Because we do not have aggregate data on website ideology for years other than 2009, this figure is based on the comScore microdata, translating units as described in section 3.2. These estimates should be taken with caution given the limitations of the comScore microdata.

There is no evidence that ideological segregation on the Internet has increased. If anything, segregation has declined as the Internet news audience has grown.

4.4 Interpretation of Magnitudes

The discussion above focused on the way Internet segregation compares to offline media and face-to-face interactions. In this section, we ask whether ideological segregation on the Internet is large or small in absolute terms.

One approach is to look at the content that liberals and conservatives encounter online. The average liberal's conservative exposure is 53 percent, similar to getting news exclusively from cnn.com. The average conservative's conservative exposure is 61 percent, similar to getting news exclusively from usatoday.com.

A second approach is to use the metaphor of online "interactions" between conservatives and liberals introduced in section 3.1. The 57 percent of Internet news consumers who are conservative are exposed to 39 percent liberals, whereas the 43 percent who are liberal are exposed to 53 percent conservatives. Therefore $0.57\,(0.39) + 0.43\,(0.53) = 45$ percent of interactions are between individuals of different ideologies. With only a single site (and therefore no segregation) this share would be $0.57\,(0.43) + 0.43(0.57) = 49$ percent. That is, the current extent of ideological segregation online decreases cross-ideology interactions by 4 percentage points, or 8 percent, relative to a benchmark of no segregation.

A third approach is to compare conservative exposure online to exposure in US states. The difference between the exposure of the average conservative and the average liberal is similar to the difference between interacting with a random resident of Minnesota or Iowa (61 percent conservative), and interacting with a a random resident of Massachusetts (52 percent conservative) or California (55 percent conservative).

A final approach is to compare the ideological segregation we estimate to previous estimates of racial segregation. As of the 2000 Census, the average isolation index for blacks across metropolitan areas was 20 unweighted, and 40 weighting by black population (Glaeser and Vigdor 2001). The average black student's school is 55 percent black; the average white student's school is 9 percent black (Orfield 2001). Racial segregation is more comparable to ideological segregation in face-to-face interactions than to ideological segregation in online news.

5 What Determines the Extent of Segregation Online?

The evidence above suggests that ideological segregation on the Internet is lower, both in absolute terms and relative to other domains of interaction, than many observers have conjectured. In this section, we discuss two key features of the economics of news markets that limit online segregation.

5.1 Vertical Differentiation and the Long Tail

Online news consumption is highly concentrated. Figure 4 shows the cumulative distribution of daily unique visits by site size. The top four sites—Yahoo! News, AOL News, msnbc.com, and cnn.com—account for more than 50 percent of all visits, the top 10 sites account for more than 60 percent, and the top 20 sites account for nearly 80 percent.

As table 2 showed, the largest sites also tend to be relatively centrist, with conservative shares close to the overall average among Internet news viewers. To reinforce this point, consider the distribution across sites of the share conservative. The unweighted distribution of site share conservative has a standard deviation of 22 percentage points and an interquartile range of 29 percentage points. Weighting by site size (average daily unique visitors), the distribution is greatly compressed. The weighted distribution has a standard deviation of 14 percentage points and an interquartile range of 7 percentage points.

Table 7 shows how segregation varies across the distribution of site size. Segregation is low within the top 10 sites (isolation index of 6.2), similar for sites ranked 11-25 (isolation of 5.8), and grows as we move to sites $26-50$ (isolation of 8.6) and $51+$ (isolation of 21.3).

Together, these facts suggest that vertical differentiation serves to limit segregation. Much of the discussion about political extremism online has focused on political blogs and other small sites. Our data shows that these sites are indeed very extreme, but they account for a negligible share of Internet news consumption.

It is beyond the scope of this paper to analyze the deeper reasons why large relatively moderate sites dominate Internet news. But the basic economics of the news business suggests that this fact should not be all that surprising. Although consumers' tastes in news are heterogeneous, they are highly correlated—most people prefer stories that are timely, well written, entertaining, and do not omit or explicitly misreport important facts. News production has high fixed costs and low marginal costs (especially online), meaning producers will be more likely to invest in creating a quality product if they can appeal to a wide audience.

It is true that the Internet allows consumers to *filter* news relatively freely, but it has not changed the fact that *reporting* or *writing* stories that are tailored to a particular point of view is costly. There is no computer program that can take a story written with liberal slant as input, and output an account of the same facts written with conservative slant. One could imagine a news site that presented the Neo-Nazi perspective on all of the day's events: first hand Neo-Nazi reports from a hurricane in Florida, a Neo-Nazi perspective on the Superbowl, and so forth. But such a site does not exist, to our knowledge, likely because the Neo-Nazi audience is too small to make such an investment worthwhile, and the preferences of Neo-Nazis for many stories are not actually all that different from those of the average consumer.

The pattern of significant vertical differentiation has remained consistent even as media technologies have changed dramatically. To the extent this flows from the underlying economics of news markets, it is likely that the pattern of low ideological segregation on the Internet will continue.

5.2 Site Segregation vs. User Segregation

As we have noted, the typical conservative's exposure to conservatives is far lower than that of an individual who gets her news exclusively from a "typical" conservative source such as foxnews.com. This is because most sites' users visit other sites as well.

Figure 5 illustrates this distinction by plotting the conservative exposure of a site's average daily visitor against the share conservative on the site (or, equivalently, the conservative exposure of an individual who gets all her news from that site). The regression line is much shallower than the 45-degree line, reflecting the fact that extreme sites are more common than extreme users. A large number of sites have share conservative greater than 80 percent or less than 40 percent. By contrast, there are no sites whose average reader has conservative exposure greater than 80 percent or less than 40 percent. Put differently, if we were to sample readers from conservative sites like drudgereport.com, we would find that most of their readers get most of their news from sites that are substantially less conservative. Similarly, if we were to sample readers from liberal sites like huffingtonpost.com, we would find that most of their readers get most of their news from sites that are substantially less liberal.

Table 8 shows cross-visiting patterns in more detail. For each of the ten most liberal and ten most conservative sites in our data, the table shows the share of their monthly visitors who visited Yahoo! News, foxnews.com, and nytimes.com in the same month. The results are striking. Visitors to the most conservative sites are typically more likely to visit nytimes.com in the same month than the average Internet user or the average visitor to Yahoo! News. Visitors to the most liberal sites are typically more likely to visit foxnews.com than the average Internet user or the average visitor to Yahoo! News.

To take an even more extreme example, visitors to stormfront.org, a “discussion board for pro-White activists and anyone else interested in White survival,” are twice as likely as visitors to Yahoo! News to visit nytimes.com in the same month. The pattern of cross-visiting contrasts with the image of online “echo chambers” where users are never exposed to opposite perspectives.

Here too, there are basic economics that drive the pattern we see. The Internet makes it easy to consume news from multiple sources. Of course many people do get news from only one source, but these tend to be light users, and their sole source tends to be one of the large relatively centrist outlets. Most of the people who visit sites like drudgereport.com or huffingtonpost.com, by contrast, are heavy Internet users, likely with a strong interest in politics. Although their political views are relatively extreme, they also tend to consume more of everything, including centrist sites and occasionally sites with conflicting ideology. Their omnivorousness outweighs their ideological extremity, preventing their overall news diet from becoming too skewed.

6 Do Conservatives and Liberals See the Same Content on the Sites They Visit?

Our segregation measure captures the extent to which liberals and conservatives visit the same outlets. We cannot observe directly whether they choose to read the same stories within those outlets. Story-level segregation could in principle be either higher or lower than outlet-level segregation.

The possibility of within-outlet sorting applies to all domains of interaction, and is in no way special to the Internet. Conservatives and liberals may both read the *Wall Street Journal*, but conservatives may prefer the editorial page while liberals prefer the international news section. An individual's neighborhood may be politically heterogeneous, but a person she seeks out to discuss politics with may be more likely than a random neighbor to share her ideology.

Note, also, that even if we could measure which stories conservatives and liberals seek out across all media, outlet-level segregation may still be an object of primary interest. On the Internet, liberals and conservatives could spend much of their time on different sections of nytimes.com, but the fact that they see the same front page, the same headlines, and the same links in the side bars might nevertheless have important benefits, as with "unexpected encounters" in traditional domains of interaction.[5] Although customization and referrals from portal pages could reduce such "unexpected encounters," at present they represent a minority approach to consuming news online.[6]

With the above caveats in mind, we ask what our data can tell us about the relationship of story-level and outlet-level segregation on the Internet. Our approach is to ask how outlet-level segregation changes on days when there is a major event that causes a spike in total news demand.

[5]"When you go to work or visit a park, it is possible that you will have a range of unexpected encounters, however fleeting or seemingly inconsequential... You cannot easily wall yourself off from contentions or conditions that you would not have sought out in advance, or that you would have avoided if you could have" (Sunstein 2001, 32). Similarly, "When you read a city newspaper or a national magazine, your eyes will come across a number of articles that you would not have selected in advance" (34).

[6]In our microdata, visits to news sites resulting from referrals by other news sites account for 13 percent of all daily visits. Among respondents to the 2008 Pew Research Center Biennial Media Consumption Survey who say they read news online, 64 percent say they never use portal pages such as iGoogle or My Yahoo! that potentially include customized news. Only 14 percent report sending a news story by e-mail in the past week, 27 percent report receiving a news story by e-mail in the past week, and 12 percent report ever receiving news items via an RSS feed. Moreover, to our knowledge, none of the major portal sites currently allow users to select news according to its political slant. The customization options typically only allow users to filter news by broad categories such as sports, crime, or local stories.

The extra consumption of conservatives and liberals on such days will presumably be devoted to reading about the event. Therefore on major news days outlet-level segregation is more representative of story-level segregation than on other days. If outlet-level segregation is normally low because liberals and conservatives can view different content on the same site, then outlet-level segregation should increase on major news days when the overlap in their story readership is higher.

We select the top news events of 2008 and 2007 as defined by the Associated Press. The top news event of 2008 is the presidential election on November 4. The top news event of 2007 is the Virginia Tech massacre on April 16.

Panel A of figure 6 shows the total number of unique visitors for all news sites in our comScore micro-data sample for each day in 2008 and 2007 respectively. In 2008, news consumption increases steadily in the weeks approaching the election, and jumps two-fold on election day itself. In 2007, there is a clear spike on the day of the shooting.

Panel B of figure 6 shows daily isolation indices estimated from the comScore microdata, using our zipcode-based ideology proxy. We rescale this measure so the mean across days is equal to the isolation index for our main measure. In 2008, we see no buildup in the weeks before the election, and no spike in segregation on election day. In 2007, we see no increase on the day of the Virginia Tech shooting. In fact, segregation on both of the major news days is actually lower than average.

Conservatives and liberals did not get their information about the top news events of 2007 and 2008 from very different sources. If anything, sources of information are less segregated when a major news event unfolds, even though such days are likely characterized by limited within-site segregation.

7 Robustness

7.1 Weighting

As discussed in section 3.3, our main segregation estimates weight users by the total number of visits they make on each medium. That is, they capture the segregation of the average visit rather than the segregation of the average user. We cannot calculate a user-weighted version of our main measure for the Internet because it is based on aggregate data. As an approximation, we use the 2008 comScore microdata to estimate that the ratio of user-weighted to visit-weighted segregation is 0.71. Applying this ratio to our main measure we estimate a user-weighted isolation index of 5.3.

7.2 Time Aggregation

Section 3.3 notes that our main segregation estimates define a visit to mean looking at a site at least once on a given day. Under reasonable assumptions, we expect the absolute magnitude of the isolation index to be higher for shorter time intervals and lower for longer time intervals.

Daily visits is the finest level of aggregation that we can compare across media. We can, however, use the 2008 comScore microdata to look at how the isolation index depends on the level of time aggregation. As in section 7.1 above, we use the ratio of user-weighted segregation in the microdata to visit-weighted segregation in our main sample to scale microdata calculations into units comparable to those of our main estimates.[7]

As noted in section 7.1 above, the user-weighted isolation index is equal to 5.3 when we define a visit to be a unique daily visit. We estimate that the isolation index falls to 3.2 when we define a visit to be a unique *monthly* visit, and increases to 9.1 and 10.8 when we define a visit to be a unique page view or a unique minute respectively. Because we do not observe offline media or face-to-face interactions at these alternative levels of aggregation, we cannot say how the relative rankings would change. The absolute magnitude of isolation for the Internet, however, remains relatively low even at the finest possible level of aggregation.

7.3 Additional Robustness Checks

We present additional robustness checks in table 9. The first row presents our baseline estimates from table 4.

The next row shows that low segregation on the Internet is not only driven by Yahoo! News and AOL News–the isolation index is still only 11.3 when these important sites are excluded.

[7] As noted in section 7.1, weighting by visits rather than users introduces some upward bias in our segregation measure. Weighting by page views or minutes increases the magnitude of this distortion, while weighting by monthly unique visits reduces it.

The following two rows present estimates for expanded sets of websites. In the first, we expand our sample to include 391 websites for which we have comScore MediaMetrix data on average daily visitors, but no PlanMetrix data on visitor ideology. For these sites, we estimate segregation using the comScore microdata and adjust the units using equation 6. We estimate that expanding the long tail of websites in this way increases the Internet isolation index from 7.5 to 9.9 percentage points.

In the next row, we compute an upper bound for the segregation we would observe if we could measure the entire population of Internet news sites. We compute the share of online news consumption accounted for by the sites in our main sample by estimating a power-law distribution for site size (Adamic 2010) and calculating the implied share of consumption accounted for by the top 119 sites (the number in our main sample). We compute an upper bound by assuming all remaining consumption is of sites with 100 percent conservative or 100 percent liberal readership. We estimate that the maximum possible value of the isolation index for the entire population of online news sites is 10.2.

The following three rows report alternative treatments of "middle of the road" respondents. Categorizing them as conservatives, categorizing them as liberal, and dropping them from the sample entirely yields isolation indices of 5.0, 7.9, and 9.1 respectively.

In the final row, we replace our conservative-liberal measure of ideology with the right-of-median zipcode ideology measure that forms our proxy in the comScore microdata, and estimate an isolation index of 1.3 for sites in both our main sample and the comScore microdata.

8 Conclusion

The evidence above suggests that ideological segregation on the Internet is low in absolute terms, higher than most offline media (excluding national newspapers), and significantly lower than segregation of face-to-face interactions in social networks. These findings may mitigate concerns expressed by Sunstein (2001) and others that the Internet will increase ideological polarization and threaten democracy.

An important caveat, however, is that none of our evidence speaks to the way people translate the content they encounter into beliefs. Both Bayesian (Gentzkow and Shapiro 2006; Acemoglu et al. 2009) and non-Bayesian (Lord et al. 1979) mechanisms may lead people with divergent political views to interpret the same information differently, and the beliefs of conservatives and liberals frequently diverge on important factual questions. That they do so despite the fact that most Americans are getting their information from the same sources emphasizes the importance of further research on the formation and evolution of beliefs.

References

Acemoglu, Daron, Victor Chernozhukov and Muhamet Yildiz. 2009. Fragility of asymptotic agreement under Bayesian learning. *Working Paper, Massachusetts Institute of Technology.*

Adamic, Lada A. 2010. Zipf, power-laws, and Pareto - a ranking tutorial. Information Dynamics Lab, HP Labs. http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html.

Adamic, Lada A. and Natalie Glance. 2005. The political blogosphere and the 2004 U.S. election: divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery,* August 21-25. Chicago, Illinois: Association of Computing Machinery, pp. 36-43.

Brooks, David. 2010. Getting Obama right. *The New York Times.* March 11. Op-Ed Section.

Campante, Filipe R. and Daniel Hojman. 2010. Media and Polarization. *Working Paper, Harvard Kennedy School.*

Carr, Nicholas. 2008. *The Big Switch: Rewiring the World, from Edison to Google.* New York: W. W. Norton & Company.

Cutler, David M., Edward L. Glaeser and Jacob L. Vigdor. 1999. The rise and decline of the American ghetto. *The Journal of Political Economy* 107, no. 3, 455-506.

DellaVigna, Stefan and Ethan Kaplan. 2007. The Fox News effect: Media bias and voting. *Quarterly Journal of Economics* 122, no. 3, 1187-1234.

DiPrete, Thomas, Andrew Gelman, Julien Teitler, Tian Zheng and Tyler McCormick. forthcoming. Segregation in social networks based on acquaintanceship and trust. *American Journal of Sociology.*

Frankel, David and Oscar Volij. 2008. Scale invariant measures of segregation. *Working paper.*

Friedman, Thomas L. 2009. Advice from grandma. *The New York Times*. November 21. Op-Ed Section.

Gentzkow, Matthew. 2006. Television and voter turnout. *The Quarterly Journal of Economics* 121, no. 3, 931-972.

Gentzkow, Matthew and Jesse M. Shapiro. 2006. Media bias and reputation. *Journal of Political Economy* 114, no. 2, 280-316.

Gentzkow, Matthew and Jesse M. Shapiro. 2008. Competition and truth in the market for news. *Journal of Economic Perspectives* 22, no. 2, 133-154.

Gentzkow, Matthew and Jesse M. Shapiro. 2010. What drives media slant? Evidence from U.S. daily newspapers. *Econometrica* 78, no. 1, 35-71.

Glaeser, Edward L. and Jacob L. Vigdor. 2001. Racial segregation in the 2000 Census: Promising news. *Center on Urban and Metropolitan Policy Survey Series.* Brookings Institution.

Glaeser, Edward L. and Bryce A. Ward. 2006. Myths and realities of American political geography. *Journal of Economic Perspectives* 20, no. 2, 119-144.

Knight, Brian G. and Chun-Fang Chiang. 2008. Media bias and influence: Evidence from newspaper endorsements. *SSRN Working Paper.*

Kohut, Andrew. 2004. The Nation: Media myopia; More news is not necessarily good news. *The New York Times.* July 11. Week in Review Section.

Leibowitz, Jon. 2010. Introductory remarks. Comments given at conference, How Will Journalism Survive the Internet Age? March 9-10, Federal Trade Commission Workshop.

Lord, Charles G., Lee Ross and Mark R. Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37, no. 11, 2098-2109.

McCarty, Nolan, Keith Poole and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge, M.A.: MIT Press.

Mullainathan, Sendhil and Andrei Shleifer. 2005. The market for news. *American Economics Review* 95, no. 4, 1031-1053.

Olken, Benjamin A. 2009. Do TV and radio destroy social capital? Evidence from Indonesian villages. *American Economic Journal: Applied Economics* 1, no. 4, 1-33.

Orfield, Gary. 2001. Schools more separate: Consequences of a decade of resegregation. *The Civil Rights Project*, University of California, Los Angeles (formerly Harvard University). http://www.civilrightsproject.ucla.edu/research/deseg/separate_schools01.php.

Prior, Markus. 2005. News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. *American Journal of Political Science* 49, no. 3, 577-592.

Sobbrio, Francesco. 2009. A citizens-editors model of news media. *Working Paper.*

Stone, Daniel F. 2010. Ideological media bias. *SSRN Working Paper.*

Stromberg, David. 2004. Radio's impact on public spending. *The Quarterly Journal of Economics* 119, no. 1, 189-221.

Stromberg, David and James Snyder. forthcoming. Press coverage and political accountability. *Journal of Political Economy.*

Sunstein, Cass R. 2001. *Republic.com.* Princeton, N.J.: Princeton University Press.

Tewksbury, David. 2005. The seeds of audience fragmentation: Specialization in the use of online news sites. *Journal of Broadcasting and Electronic Media* 49, no. 3, 332–348.

Von Drehle, David. 2004. Political split is pervasive: Clash of cultures is driven by targeted appeals and reinforced by geography. *The Washington Post.* April 25. A01.

White, Michael J. 1986. Segregation and diversity measures in population distribution. *Population Index* 52, no. 2, 198-221.

Appendices

A Alternative Segregation Measures

We present estimates for two alternative measures of ideological segregation: the dissimilarity index (Cutler et al. 1999), and the symmetric Atkinson index (Frankel and Volij 2008). We also show how our main results change when we do not implement the small-sample correction discussed in section 3.2.

The dissimilarity index is defined as:

$$S_m^D = \frac{1}{2}\sum_{j\in J_m}\left|\frac{cons_j}{cons_m} - \frac{lib_j}{lib_m}\right|.$$

It can be interpreted as the share of conservative (or liberal) visits that would need to be redistributed across media for the share conservative to be uniform across outlets.

The symmetric Atkinson index is defined as:

$$S_m^A = 1 - \sum_{j\in J_m}\left(\frac{lib_j}{lib_m}\right)^{\frac{1}{2}}\left(\frac{cons_j}{cons_m}\right)^{\frac{1}{2}}.$$

Frankel and Volij (2008) shows that the ordering defined by this index is the unique measure of segregation satisfying a set of intuitive axioms, including scale invariance.

Appendix table 1 presents the results. Because social networks do not partition the set of individuals, we cannot compute these indices for the acquaintance groups defined in the GSS. The first column repeats our main results from table 4. The second column shows how the results change when we do not implement a finite-sample correction. The qualitative pattern is if anything strengthened, with segregation increasing as expected, especially for zipcodes and local newspapers where we have very small samples for each "outlet." This column is the most relevant benchmark to compare to the alternative indices, since they are not corrected for finite-sample bias. The final two columns show that the qualitative pattern for the dissimilarity index and the symmetric Atkinson index is similar to that for the unadjusted isolation index. The only notable change is that cable news looks more segregated than the Internet on the dissimilarity measure.

Table 1: Size and Share Conservative of Major News Media

US adult population: 67 percent conservative

Medium	Share Conservative	Share of Daily Visits
Cable	.70	.31
Local newspapers	.68	.26
Broadcast news	.67	.26
Magazines	.58	.03
Internet	.57	.11
National newspapers	.57	.04

Note: Share of daily visits is the ratio of the sum of average daily unique visitors across all outlets in the medium to the sum of average daily unique visitors across all outlets in all media. Share conservative is the average of share conservative across outlets, weighting each outlet in the medium by its average daily unique visitors. Internet data are from comScore; data on other media are from MRI.

Table 2: Size and Share Conservative of Online News Outlets

Ten Largest			Most Conservative			Most Liberal		
Site	Share Conservative	Daily UV ('000)	Site	Share Conservative	Daily UV ('000)	Site	Share Conservative	Daily UV ('000)
drudgereport.com	.93	475	billoreilly.com	1.00	10	thinkprogress.org	.06	12
foxnews.com	.88	1,159	rushlimbaugh.com	.99	43	blogcritics.org	.17	17
AOL News	.62	3,971	glennbeck.com	.98	38	bvblackspin.com	.17	57
usatoday.com	.60	518	humanevents.com	.97	21	moveon.org	.19	21
msnbc.com	.57	3,264	townhall.com	.96	42	BBC News	.22	472
Yahoo! News	.55	6,455	thestate.com	.94	36	blogtalkradio.com	.22	33
cnn.com	.54	2,650	aclj.org	.93	18	reddit.com	.23	36
nytimes.com	.40	879	cnsnews.com	.93	12	newsvine.com	.25	56
huffingtonpost.com	.30	583	drudgereport.com	.93	475	alternet.org	.26	16
BBC News	.22	472	realclearpolitics.com	.93	41	dailykos.com	.27	26

Notes: Average daily unique visitors is reported in 1000s. Data are from comScore. See section 2 for details on the construction of size and share conservative measures. To improve precision, sites with fewer than 10000 average daily unique visitors are excluded from "most conservative" and "most liberal" lists.

Table 3: Size and Share Conservative of Offline News Outlets

Magazines

Magazine	Share Conservative	Market Share
Barron's	.70	.02
BusinessWeek	.67	.07
U.S. News & World Report	.65	.15
Fortune	.63	.03
Forbes	.63	.04
TIME	.58	.32
Newsweek	.56	.28
The Economist	.42	.03
The New Yorker	.30	.06
The Atlantic	.26	.00

National Newspapers

Paper	Share Conservative	Market Share
USA Today	.68	.40
The Wall Street Journal	.68	.29
The New York Times	.33	.31

Broadcast News

Channel	Share Conservative	Market Share
CBS	.70	.28
NBC	.69	.29
ABC	.69	.31
BBC	.55	.06
PBS	.47	.07

Cable

Channel	Share Conservative	Market Share
Fox News Channel	.81	.36
Bloomberg Television	.74	.01
CNBC	.66	.13
CNN	.64	.33
MSNBC	.62	.17

Data are from MRI. See section 2 for details on the construction of size and share conservative measures. Market share is the ratio of the outlet's daily readers/viewers to the sum of daily readers/viewers across all listed outlets in the medium. Market shares may not sum to one due to rounding.

Table 4: Ideological Segregation by Medium and Type of Interaction

	Conservative Exposure of		
	Conservatives	Liberals	Isolation Index
Internet	.606	.531	.075
Offline Media			
Broadcast News	.677	.660	.018
Magazines	.588	.558	.029
Cable	.712	.679	.033
Local Newspapers	.685	.644	.041
National Newspapers	.612	.508	.104
Face-to-Face Interactions			
County	.682	.622	.059
Zipcode	.637	.543	.094
Voluntary Associations	.625	.480	.145
Work	.596	.428	.168
Neighborhood	.627	.439	.187
Family	.690	.447	.243
People You Trust	.675	.372	.303

Notes: Internet data are from comScore. County, zipcode, and offline media data are from MRI. Voluntary associations, work, neighborhood, family, and "people you trust" data are from the GSS. See section 3 for details on the construction of exposure and isolation measures.

Table 5: Exposure by Detailed Ideology

	Exposure to:				
Exposure of:	Very Liberal	Somewhat Liberal	Middle of the Road	Somewhat Conservative	Very Conservative
Very Liberal	0.130	0.186	0.345	0.192	0.148
Somewhat Liberal	0.112	0.190	0.357	0.191	0.150
Middle of the Road	0.100	0.172	0.377	0.199	0.152
Somewhat Conservative	0.097	0.161	0.347	0.214	0.182
Very Conservative	0.087	0.147	0.309	0.212	0.246
All Internet Users	0.102	0.170	0.352	0.202	0.174

Notes: Data are from comScore. See section 3 for definition of exposure.

Table 6: Trends in Ideological Segregation Online

	Cons. Exposure of Conservatives	Cons. Exposure of Liberals	Isolation Index
2004	.635	.492	.143
2006	.625	.506	.118
2007	.625	.505	.120
2009	.606	.531	.075

Notes: Data are from comScore microdata, with estimates rescaled to match the 2009 isolation index reported in table 4. See section 3 for details.

Table 7: Ideological Segregation by Site Size

Subset of Sites with Size Rank	Share of Daily Visitors	Cons. Exposure of Conservatives	Cons. Exposure of Liberals	Isolation Index
1-10	.687	.599	.536	.062
11-25	.147	.584	.526	.058
26-50	.094	.610	.525	.086
51+	.065	.695	.482	.213

Notes: Data are from comScore. Daily visitors is the sum of average daily unique visitors across all sites. See section 3 for details on the construction of exposure and isolation measures.

Table 8: Cross-Visiting Online

	Share Visiting in the Same Month		
Site	Yahoo! News	foxnews.com	nytimes.com
Monthly Visitors of:			
Any Internet Site	.24	.05	.06
Yahoo! News	1.00	.09	.12
Most Conservative			
billoreilly.com	.38	.50	.22
rushlimbaugh.com	.50	.49	.31
glennbeck.com	.44	.44	.21
humanevents.com	.51	.44	.34
townhall.com	.51	.42	.33
thestate.com	.43	.28	.21
aclj.org	.42	.25	.15
cnsnews.com	.61	.60	.44
drudgereport.com	.52	.44	.30
realclearpolitics.com	.60	.53	.51
Most Liberal			
thinkprogress.org	.57	.33	.48
blogcritics.org	.30	.13	.21
bvblackspin.com	.25	.12	.14
moveon.org	.41	.12	.27
BBC News	.39	.18	.25
blogtalkradio.com	.24	.07	.14
reddit.com	.35	.12	.28
newsvine.com	.37	.24	.21
alternet.org	.45	.24	.40
dailykos.com	.45	.24	.40

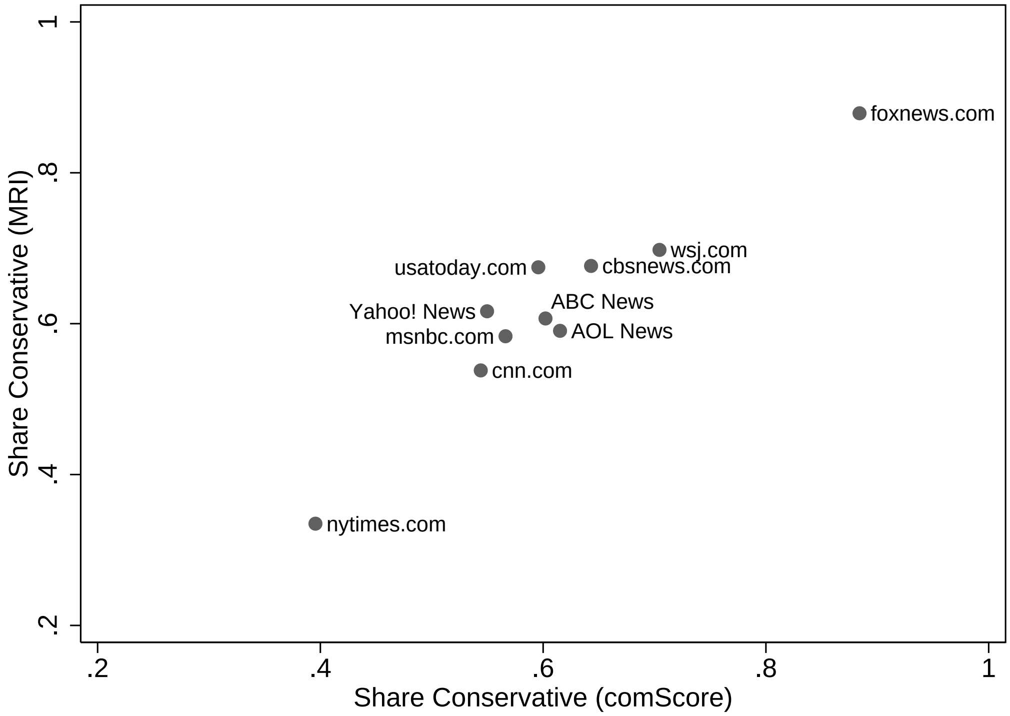
Notes: Data are from comScore. Rows list the share of monthly visitors of a given site that visit Yahoo! News, foxnews.com, and nytimes.com, respectively, in the same month, averaged over months in 2009. To improve precision, sites with fewer than 10000 average daily unique visitors are excluded from “most conservative” and “most liberal” lists.

Table 9: Robustness Checks

	Share of Interactions w/Conservatives (Internet)		
	Conservative	Liberal	Isolation Index
Baseline	.606	.531	.075
Exclude AOL & Yahoo!	.622	.509	.113
Expand the Set of News Sites			
391 Websites in comScore Microdata	.616	.517	.099
All News Websites Upper Bound	.617	.516	.102
Treat “Middle of the Road” as Conservatives	.742	.692	.050
Treat “Middle of the Road” as Liberals	.425	.346	.079
Drop “Middle of the Road”	.618	.528	.091
Right-of-Median Zipcode as Ideology Measure	.510	.497	.013

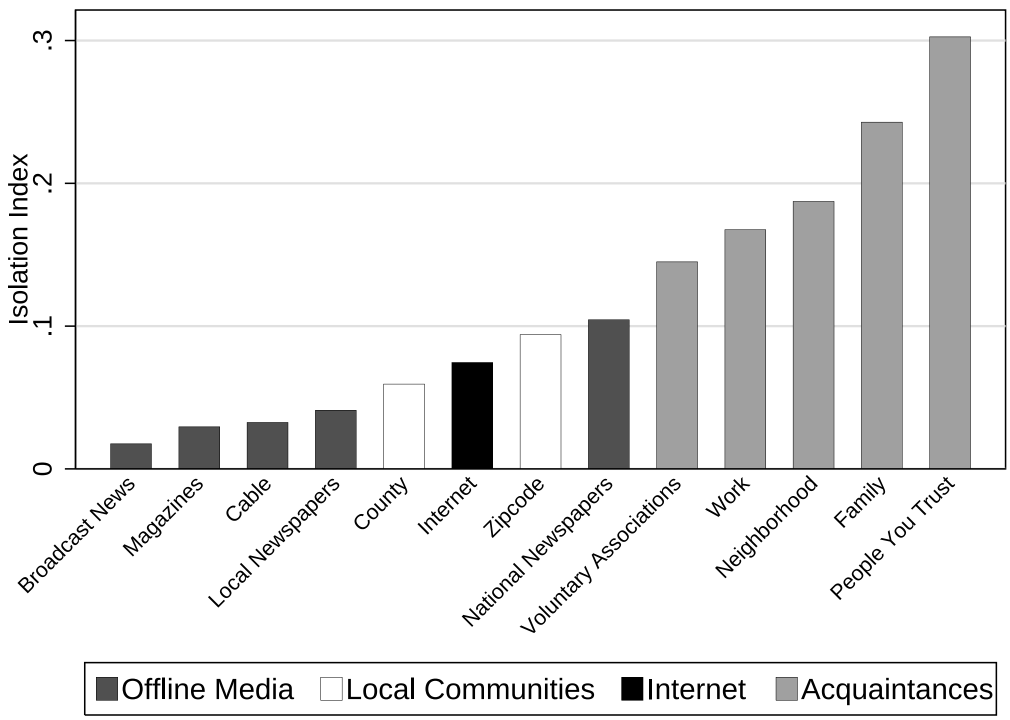
Notes: Data are from comScore. See section 3 for details on the construction of exposure and isolation measures. Zipcode ideology measure is constructed from Federal Election Commission data on political contributions. See section 7 for details.

Figure 1: Comparison of MRI and comScore Share Conservative



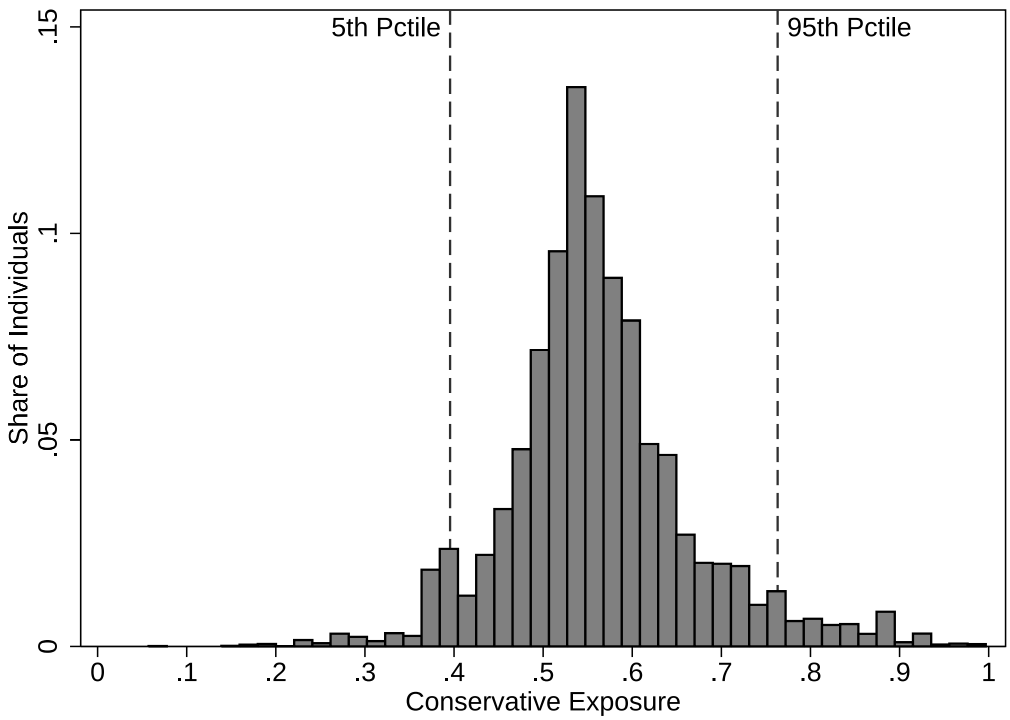
Notes: Data are from comScore and MRI. Share conservative is the estimated share of daily visitors who are conservative. See section 2 for details on variable construction.

Figure 2: Ideological Segregation by Medium and Type of Interaction



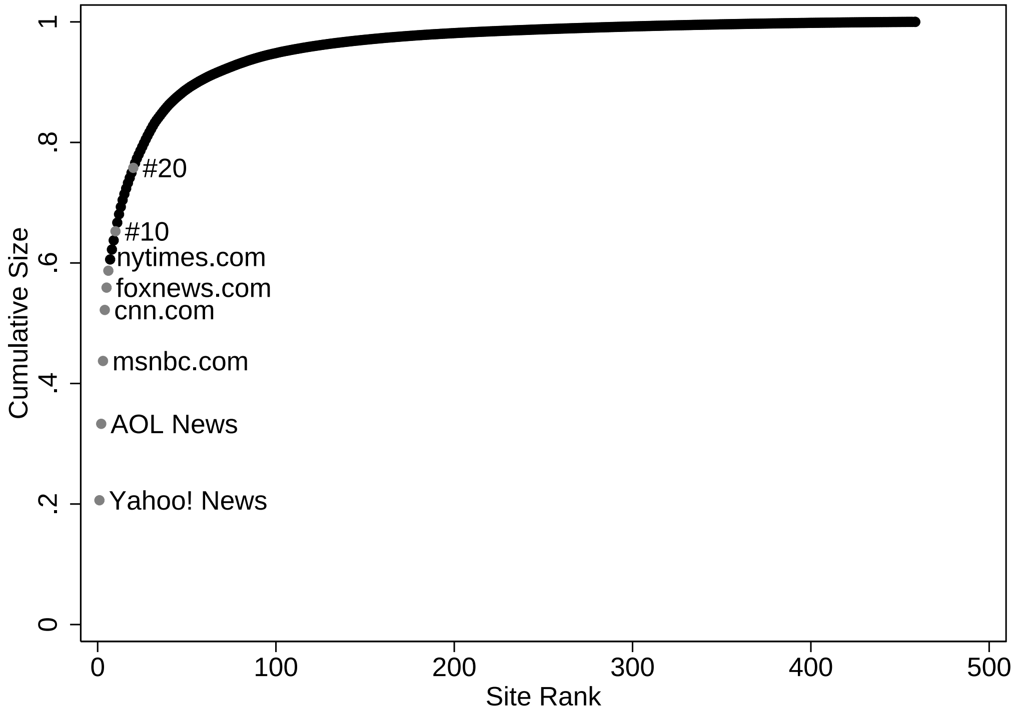
Notes: Internet data are from comScore. County, zipcode, and offline media data are from MRI. Voluntary associations, work, neighborhood, family, and "people you trust" data are from the GSS. See section 3 for details on the construction of the isolation index.

Figure 3: Distribution of Conservative Exposure across Internet Users



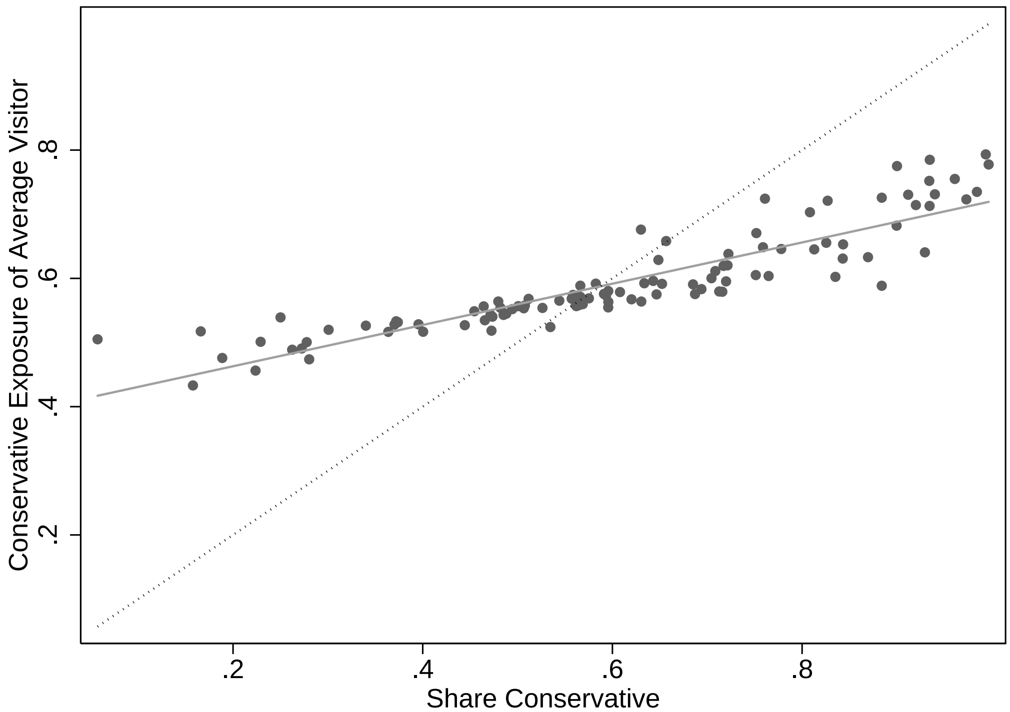
Notes: Data are from comScore. See section 3 for details on the construction of the exposure index.

Figure 4: Cumulative Distribution of Internet Unique Visits



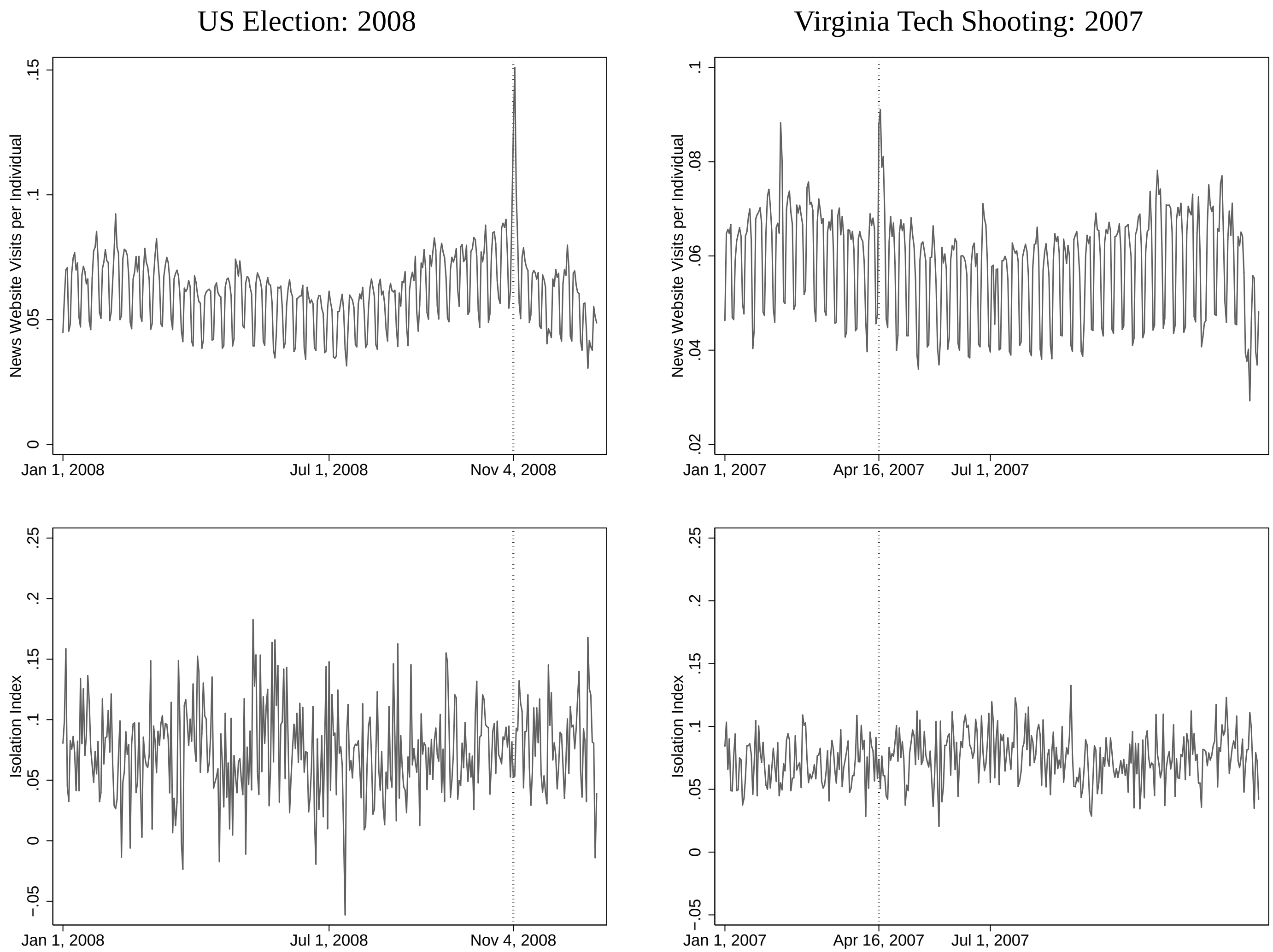
Notes: Data are from comScore. Size is measured by average daily unique visitors.

Figure 5: Visitor Exposure vs. Site Share Conservative



Notes: Data are from comScore. Figure plots conservative exposure of average daily visitor against the share of daily visitors who are conservative. The solid line is an OLS regression fit; the dotted line is the 45-degree line. See section 3 for details on the construction of the exposure index.

Figure 6: Online Daily Visitors and Segregation by Day



Notes: Data are from comScore microdata. In top panel, news website visits per individual is the average across individuals of the number of news websites in our main sample visited on each day. In bottom panel, the isolation index is scaled so that its mean across days is equal to the isolation index for the Internet in table 4.

Appendix Table 1: Alternative Segregation Measures

	Isolation Index	Isolation Index (Unadjusted)	Dissimilarity Index	Atkinson Index
Internet	.075	.079	.184	.048
Offline Media				
Broadcast News	.018	.019	.093	.010
Magazines	.029	.031	.107	.016
Cable	.033	.033	.190	.021
Local Newspapers	.041	.128	.300	.101
National Newspapers	.104	.109	.309	.056
Face-to-Face Interactions				
County	.059	.129	.297	.098
Zipcode	.094	.416	.564	.379

Notes: Internet data are from comScore. County, zipcode, and offline media data are from MRI. See section 3 for details on construction of isolation index. See appendix A for definitions of unadjusted isolation index, dissimilarity index, and Atkinson index.