Ideological Segregation Online and Offline

Matthew Gentzkow Jesse M. Shapiro*
Chicago Booth and NBER

First Version: April 5, 2010
This Version: March 18, 2011

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.

Keywords: news, internet, echo chambers
JEL: D83, L86

*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 Jim Webster, Henry Farrell, Kelly Garrett, Markus Prior, and numerous seminar participants 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.

1 Introduction

Democracy is most effective when citizens have accurate beliefs (Becker 1958; Downs 1957). 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

[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).

conservative exposure to be the average share conservative on the outlets she visits. For example, if 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' acquaintances and political discussants as reported in the 2006 General Social Survey and the 1992 Cross-National Election Study.

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 news (1.8), cable television news (3.3), magazines (4.7), and local newspapers (4.8), 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 5.1 to 4.1.

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) and political discussants (39.4).

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, most online news consumption is 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 present a series of robustness checks. We also consider the possibility that segregation at the level of individual stories may differ from segregation at the level of the news outlet, and present several pieces of evidence suggesting that story-level segregation is unlikely to be very different from the outlet-level segregation we measure.

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. Apart from Lawrence et al.'s (2010) analysis of the ideological polarization of blog audiences, most evidence on ideological segregation online comes from data on content or link structures rather than consumption (e.g., Adamic and Glance 2005; Hargittai et al 2008).[2]

A large literature considers the causes and effects of political polarization (McCarty et al 2006; Glaeser and Ward 2006), which Prior (2008), Campante and Hojman (2010), and others 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 2010), political beliefs and behavior (Prior 2005, Gentzkow 2006, DellaVigna and Kaplan 2007, Knight and Chiang 2008), and social capital (Olken 2009). A related literature considers whether news consumers are motivated by information-seeking or a desire for reinforcement (DiMaggio and Sato 2003, Mullainathan and Shleifer 2005, Gentzkow and Shapiro 2006, Bennett and Iyengar 2008, Garrett 2009a and 2009b, Iyengar and Hahn 2009). A separate literature in economics considers the effects of the Internet on communication more broadly (Glaeser 1998, Rosenblat and Mobius 2004).

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. Section 6 presents robustness checks and an analysis of the segregation of content (as opposed to site) viewership. Section 7 concludes.

[2]Benkler and Shaw (2010) compare characteristics such as the extent of user participation between right-wing and left-wing blogs. Baum and Groeling (2008) argue that online sources engage in more partisan filtering of content than news wires. Tewksbury (2005) presents evidence on demographic (not specifically ideological) specialization in online news audiences. Webster (2005) compares the fragmentation of broadcast and cable television network audiences. Stroud (2008) uses data from the 2004 National Annenberg Election Survey to compare the correlation between ideology and outlet choice across several media types.

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.[3] 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 Media Metrix. Media Metrix 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 ideology using data from comScore Plan Metrix. Plan Metrix 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]". The average number of daily unique visitors in each category is reported by comScore for each site for each month. We average these figures over the twelve months in 2009. We refer to those who

[3]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).

report being "middle of the road" in this and other datasets as "moderates."

We use the "political outlook" measure of ideology because it is directly comparable to the measure available in our source for offline media. In section 6.1.3 we show that estimated segregation is essentially unchanged when we measure ideology using party affiliation.

Plan Metrix 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.

We also 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 per year and contain the domain name of each site visited. We match sites in this data to our set of 119 Plan Metrix sites.

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. (See the online appendix for a complete list of sites in the comScore microdata.) 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 and 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 come from questions asking the number of hours respondents viewed CNN, Fox News, MSNBC, CNBC, and Bloomberg cable networks respectively in the last 7 days.

If the number of hours viewed is less than or equal to 7, then we assume that the number of days in the last 7 on which the respondent viewed the network is equal to the number of hours viewed. If the number of hours viewed is greater than 7, we assume that the respondent viewed the network on all of the last 7 days.

Data on broadcast television come 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 the number of days in the most recent publication period on which the respondent read *The Atlantic*, *Barron's*, *BusinessWeek*, *The Economist*, *Forbes*, *Fortune*, *The New Yorker*, *Newsweek*, *Time*, and *U.S. News & World Report* respectively.

Data on local newspapers come from a free response question asking which newspapers the respondent read in the last 24 hours. We code a respondent as reading a local newspaper if she read a daily newspaper 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.

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) and from the US module of the 1992 Cross-National Election Study (CNES).

The 2006 wave of the GSS (Davis and Smith 2009) 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 define the share who are conservative in each group to be the number the respondent identifies as strongly conservative divided by the number identified as either strongly conservative or strongly liberal.

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.

Data on respondents' political ideology come 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 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.

The CNES (Beck et al. 2000) measures political communication during a presidential election. Beck et al. (2002) use it to measure the relationship between an individual's ideology and that of her social network. A total of 1,318 respondents were asked to report their political ideology and to list up to four people with whom they discussed "important matters" in the last six months, with

the option to add a fifth person "you talked with most about the events of the recent presidential election campaign." These "discussants" were then contacted directly and asked to report their own political ideologies.

Respondents and discussants report their political ideology in response to the question "Many people use the terms 'liberal' and 'conservative' to recognize different political opinions. I have a scale that runs from 1 to 10, where 1 is the most liberal position and 10 is the most conservative positive. Using any number between 1 and 10, where you do place yourself on this scale when you think of your own political views?"

In our main analysis we consider only "political discussants": those with whom respondents report discussing politics "sometimes" or "often." As we show in the online appendix, when we include all respondents the estimated isolation index is close in magnitude to the "people the respondent trusts" category in the GSS. Because the GSS data use respondents' perceptions of acquaintance ideology and the CNES data use discussants' self-reported ideology, the reasonably close agreement between these two estimates provides some comfort that the GSS estimates are not severely distorted by respondents' biased perceptions (Fowler et. al. forthcoming).

The CNES is self-weighting so we do not employ sampling weights. In the online appendix we show that our results are robust to dropping "topcoded" respondents who list 5 political discussants.

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 ($\rho > 0.9$).

As we show in section 3.1 below, our measure of segregation depends only on the size and ideological composition of news outlets. The high level of agreement on these two aggregates

between MRI (self-reported media consumption) and comScore (measured media consumption) therefore provides some confidence in the accuracy of our segregation measures even for domains where only self-reported media consumption data are available (Prior 2009).

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. Let I_{lib} and I_{cons} represent the sets of liberals and conservatives respectively. Each i is in either I_{lib} or I_{cons}.

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

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

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

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 summation, $\sum_{j \in J_m} \left(\frac{cons_j}{cons_m} \cdot \frac{cons_j}{visits_j} \right)$, is then the visit-weighted average exposure of conservatives. S_m is equal to the average conservative exposure of conservatives minus the average conservative exposure of liberals.

The isolation index captures the extent to which conservatives disproportionately visit outlets whose other visitors are conservative. The index 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). With "liberals watching and reading mostly or only liberals" (Sunstein 2001, 4-5), and conservatives behaving analogously, S_m would be close to one.

To the extent that the content of a news outlet is related to the composition of its audience (Gentzkow and Shapiro 2010), the isolation index can also be viewed as a proxy for the extent to which liberals and conservatives are exposed to different facts and opinions.

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

3.2 Estimation

We estimate isolation for each medium using an appropriately defined sample analogue of equation (1). For each medium, we compute the standard error of the estimate using a bootstrap. We report standard errors (which are small enough that they do not affect the comparisons we make) in the online appendix.

We estimate conservative exposure for each Internet user in the 2008 comScore microdata. We define an individual's conservative exposure in a given year to be the average estimated share conservative of the sites they visit weighted by the number of days in the year on which they made at least one visit.

Appendix A presents the details of our calculations. Here, we highlight three important conceptual issues that arise in measuring isolation in our data.

First, we treat ideology as binary, with all respondents having a true ideology which is either liberal or conservative. We impute the unobserved ideology of moderates by assuming that the share conservative among moderates who visit a given outlet is equal to the share conservative among visitors to the outlet who declare an ideology. This approach will tend to overstate the extent of segregation if, as seems likely, those who describe themselves as moderate have less strongly held political views, and therefore less ideologically segregated news consumption patterns, than those who declare an ideology. In appendix B, we argue using auxiliary data that our assumption about the ideology of moderates is plausible. In section 6.1.3, we present segregation measures

that use different assumptions to impute the ideology of moderates. In the online appendix, we present estimates of segregation using an ordinal generalization of the isolation index that does not require us to classify respondents as liberal or conservative.

Second, the index we calculate 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 6.1.1, we use the comScore microdata to estimate the segregation of Internet users and compare it to the segregation of Internet visits.

Third, 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 6.1.2, we argue that our conclusions are robust to using coarser or finer levels of time aggregation.

4 Main Results

4.1 Segregation Online and Offline

In table 1, we report the breakdown of reported ideology for US adults and the different media in our sample. In the MRI survey, 42 percent of adults describe themselves as very or somewhat conservative, 21 percent describe themselves as very or somewhat liberal, and the rest describe themselves as moderate. (Note that self-described conservatives outnumber self-described liberals in both the GSS and the National Election Study; see National Opinion Research Center 2009 and American National Election Studies 2009, respectively.) The ideological compositions of different

media are fairly 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 size and ideological composition of 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. The most conservative sites (according to the ratio of conservative to liberal daily visitors) are billoreilly.com, rushlimbaugh.com, and glennbeck.com, all personal sites of conservative radio or television hosts. The most liberal sites are thinkprogress.org (a liberal blog), blogcritics.org (a blog and news aggregation site), and bvblackspin.com (a blog hosted on AOL's Black Voices site).

Table 3 shows the size and ideological composition of 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. Thirteen percent of Fox News' audience is liberal, and 26 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 estimated conservative exposure of conservatives on the Internet is 60.6 percent. The estimated 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 cable (3.3) and magazines (4.7), then local newspapers (4.8), 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 5.1. Holding the distribution of offline media consumption constant, we estimate that removing the Internet would reduce this number to 4.1.

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) and political discussants (39.4).

Figure 2 shows the same estimates in a different way. Ideological segregation on the Internet is 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.

Figure 3 plots the distribution of conservative exposure across individuals in 2008. 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 conservative exposure of 88 percent, putting her at the 99th percentile. Someone who gets all her news from nytimes.com has 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 2.) 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

Figure 4 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, with estimates scaled relative to the year 2008. 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. Our exploration of the data suggests that the decline between 2007 and 2008 is attributable to a moderation in the audience of several very conservative sites.

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. Suppose, hypothetically, that each visitor to an Internet news outlet interacts with one randomly chosen other visitor to the same outlet. 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 (share conservative = 61 percent), and interacting with a random resident of Massachusetts (share conservative = 52 percent) or California (share conservative = 55 percent). For reference, in the 2008 presidential election McCain won 45 percent of the two-party vote in Minnesota and Iowa, as against 37 and 38 percent in Massachusetts and California, respectively (National Archives 2008).

5 What Determines the Extent of Segregation Online?

The facts presented so far suggest that ideological segregation on the Internet is lower, both in absolute terms and relative to other domains of interaction, than many observers have conjectured. We highlight two features of the economics of news markets that potentially limit online segregation.

First, online news sites are vertically differentiated, in the sense that a large amount of traffic goes to a small number of mainstream news sites that, at least by revealed preference, are considered high quality by most consumers. Much of the discussion about political extremism online has focused on political blogs and other small sites. Our data show that some of these sites are indeed very extreme, but they account for a negligible share of Internet news consumption. Most consumption is instead concentrated in a small number of centrist sites.

Figure 5 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. To illustrate the fact that these large sites are relatively centrist, consider the distribution across sites of 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 6 shows that the isolation index is much greater for the smallest sites in the sample than for the largest.

Second, users are not restricted to get all their news from one site. The typical conservative or liberal *site* is therefore far more extreme than the diet of the typical conservative or liberal *user*.

Figure 6 illustrates this distinction by plotting the conservative exposure of a site's average

daily visitor against the estimated 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 7 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. 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. Consistent with these facts, we show in the online appendix that many of the most ideologically extreme sites have an unusually high share of visitors who report being actively involved in politics.

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.

6 Additional Results

6.1 Robustness

6.1.1 Weighting

As discussed in section 3.2, 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 percentage points.

6.1.2 Time Aggregation

Section 3.2 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 6.1.1 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.[4]

As noted in section 6.1.1 above, the user-weighted isolation index is equal to 5.3 percentage points when we define a visit to be a unique daily visit. We estimate that the user-weighted isolation index falls to 3.2 percentage points when we define a visit to be a unique *monthly* visit, and increases to 9.1 and 10.8 percentage points 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.

6.1.3 Other Robustness Checks

We present additional robustness checks in table 8. 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

[4]As noted in section 6.1.1, 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.

and AOL News–the isolation index is still only 11.3 percentage points when these important sites are excluded.

The following three rows present estimates for expanded sets of websites. First, we add Google News to our sample. (Google News is excluded from our main sample because comScore classifies it as a search site rather than a news site.) Adding this site reduces the Internet isolation index from 7.5 to 7.2 percentage points.

Next, we expand our sample to include 391 websites for which we have comScore Media Metrix data on average daily visitors, but no Plan Metrix data on visitor ideology. For these sites, we estimate segregation using the comScore microdata and rescale the units so that the estimates agree for the set of overlapping sites. 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. The sites in this sample are listed in the online appendix.

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 percentage points.

The following five rows report alternative treatments of moderate 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 percentage points respectively. Assuming that the share conservative among moderates on each site is equal to the overall share conservative on the Internet yields an isolation index of 4.7 percentage points . Assuming that moderates on all sites are half conservative and half liberal yields an isolation index of 4.8 percentage points. In the online appendix, we present results for other media and for face-to-face interactions using the latter two alternative assumptions.

The following two rows report isolation measures replacing our conservative-liberal measure of ideology with alternative ideology measures. First, we use a measure of

party affiliation. Plan Metrix respondents are asked "Do you consider yourself to be a...[Republican/Democrat/Independent/Other/No affiliation]?" We classify Republicans as conservative and Democrats as liberal, treating all other respondents in parallel with our treatment of moderate respondents in our main analysis. The size-weighted correlation between our main measure of share conservative and the party-based measure is 0.89, and in the table we show that the isolation index goes down slightly from 7.5 to 7.1 percentage points using the party-based measure.

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

6.2 Outlet-level vs. Content-level Segregation

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. The possibility of within-outlet sorting applies to all media–newspapers consist of many articles, most of which are not read by most readers. In addition, outlet-level segregation per se is of interest, because it determines the extent to which liberals and conservatives are exposed to the same front page, side-bar links, and headlines as they locate their preferred content.[5]

With those caveats in mind, we explore story-level segregation on the Internet by asking how outlet-level segregation changes on days when there is a major event that causes a spike in total news demand. 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,

[5]Although customization and referrals from portal pages could reduce such "unexpected encounters," at present they represent a minority approach to consuming news online. 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 (Pew Research Center for the People and the Press 2008). 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.

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 (Crary 2007, Star News Online 2008). 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.

The top two panels of figure 7 show 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.

The bottom two panels of figure 7 show 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.

6.3 Non-ideological Segregation

To place our results in the context of other forms of segregation in US society, figure 8 presents the isolation index for race, gender, education, and income for online media, offline media, geographic location, and political discussants. The format parallels that of figure 2.

The figure exhibits the familiar and striking pattern of racial geographic segregation. The racial isolation index for US zipcodes is 49.1 percentage points and for counties is 21.4 percentage points. The isolation index for political discussants is even higher (81.9). Local newspapers—whose segregation tends to track that of metropolitan areas—have a racial isolation index of 12.8 percentage points. Other news media, including Internet news, have low levels of racial segregation.

Geographic isolation by education is less severe than by race. Dividing households into those with a college graduate and those without, we compute an isolation index of 21.4 percentage points for zipcodes and 8.6 percentage points for counties. The isolation index for political discussants is again higher (41.9). The most segregated news medium by education is national newspapers (8.2), followed by local newspapers (5.9) and magazines (3.0). The corresponding education isolation index for Internet news is 1.7 percentage points.

The pattern of segregation by income is broadly similar to the pattern of segregation by education. The Internet isolation index for income is 0.3 percentage points, and the most segregated news medium is local newspapers (9.4).

Not surprisingly, the genders are not very segregated geographically. Indeed, segregation by gender is generally quite low. The most segregated news media by gender are the Internet (3.5) and magazines (1.2).

7 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. Internet news consumers with homogeneous news diets are rare. These findings may mitigate concerns expressed by Sunstein (2001) and others that the Internet will increase ideological polarization and threaten democracy.

We trace our findings back to two key properties of Internet news demand: (i) news sites are highly vertically differentiated, and (ii) news consumers visit multiple sites. We take both properties as given for the purposes of the analysis in the paper, but both flow from the fundamental economics of the news media.

Consider first the fact that large and relatively moderate sites dominate Internet news. 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.

Consider next the tendency of news consumers to visit multiple outlets, and the related fact that even visitors to ideologically extreme sites have fairly moderate news diets. 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 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. These patterns accord with evidence on "long tail" consumption in other domains, such as movie rentals (Elberse 2008).

If we are correct in attributing our findings to these deeper economic forces, then we can have some confidence that the pattern of low segregation online will continue as the Internet news market develops.

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. Accessed at http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html on September, 2010.

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.

American National Election Studies. 2009. *Advance release of the 2008-2009 ANES Panel Study: Frequency Codebook.* Version 20090616 (June 16). Accessed at http://www.electionstudies.org/ studypages/2008_2009panel/anes2008_2009panel_fcodebook.txt on August 30, 2010.

Baum, Matthew A. and Tim Groeling. 2008. New media and the polarization of American political discourse. *Political Communication* 25: 345-365.

Beck, Paul Allen, Russell J. Dalton and Robert Huckfeldt. 2000. Cross-national election studies: United States study, 1992. *ICPSR Study No. 6541*. Accessed at icpsr.org on April 9, 2010.

Beck, Paul Allen, Russell J. Dalton, Steven Greene and Robert Huckfeldt. 2002. The social calculus of voting: Interpersonal, media, and organizational influences on presidential choices. *American Political Science Review* 96, no. 1, 57-73.

Becker, Gary S. 1958. Competition and democracy. *Journal of Law and Economics* 1, 105-109.

Benkler, Yochai and Aaron Shaw. 2010. A tale of two blogospheres: Discursive practices on the left and right. Berkman Center Working Paper, March 2010.

Bennett, W. Lance and Shanto Iyengar. 2008. A new era of minimal effects? The changing foundations of political communication. *Journal of Communication* 58, no. 4, 707-731.

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.

Carrington, William J. and Kenneth R. Troske. On measuring segregation in samples with small units. *Journal of Business and Economic Statistics* 15, no. 4, 402-409.

Crary, David. 2007. Massacre at Virginia Tech voted top news story of 2007. Associated Press (December 20).

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.

Davis, James Allan and Tom W. Smith. 2009. *General Social Surveys 1972-2008.* Chicago, IL: National Opinion Research Center.

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

DiMaggio, Paul and Kyoko Sato. 2003. Does the Internet Balkanize political attention? A test of the Sunstein thesis. *Paper presented at the Annual Meeting of the American Sociological Association, Atlanta, GA August 16 2003.*

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.*

Downs, Anthony. 1957. An economic theory of political action in a democracy. *Journal of Political Economy* 65, no. 2, 135-150.

Elberse, Anita. 2008. Should you invest in the long tail? *Harvard Business Review* July-August 2008, 88-96.

Fowler, James H., Michael T. Heaney, David W. Nickerson, John F. Padgett and Betsy Sinclair. forthcoming. Causality in political networks. *American Politics Research*.

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.

Garrett, R. Kelly. 2009a. Echo chambers online? Politically motivated selective exposure among Internet news users. *Journal of Computer-mediated Communication* 14, 265-289.

Garrett, R. Kelly. 2009b. Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication* 59, 676-699.

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. 1998. Are cities dying? *Journal of Economic Perspectives* 12, no. 2, 139-160.

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

Hargittai, Eszter, Jason Gallo and Matthew Kane. 2008. Cross-ideological discussions among conservative and liberal bloggers. *Public Choice* 134, no. 1-2, 67-86.

Iyengar, Shanto and Kyu S. Hahn. 2009. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication* 59, 19-39.

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.

Lawrence, Eric, Jon Sides and Henry Farrell. 2010. Self-segregation or deliberation? Blog readership, participation and polarization in American politics. *Perspectives on Politics* 8, no. 1, 141-157.

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.

National Archives. 2008. *2008 Presidential Election Popular Vote Totals*. Acccessed at http://www.archives.gov/federal-register/electoral-college/2008/popular-vote.html on February 15, 2011.

National Opinion Research Center. 2009. *General Social Surveys 1972-2008: Cumulative Codebook*. June 2009. Accessed at http://publicdata.norc.org:41000/gss/documents//BOOK/2008%20GSS%20Codebook.pdf on August 30, 2010.

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

Pew Research Center for the People and the Press. 2008. *Biennial Media Consumption Survey 2008 Final Topline*. Accessed at http://people-press.org/reports/questionnaires/444.pdf on March 23, 2010.

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.

Prior, Markus. 2008. Are hyperlinks "weak ties"? in Joseph Turow and Lokman Tsui, Eds., *The Hyperlinked Society: Questioning Connections in the Digital Age.* Ann Arbor, M.I.: University of Michigan Press.

Prior, Markus. 2009. Improving media effects research through better measurement of news exposure. *Journal of Politics* 71, no. 3, 893-908.

Ransom, Michael R. 2000. Sampling distributions of segregation indexes. *Sociological Methods and Research* 28, no. 4, 454-475.

Rosenblat, Tanya and Markus M. Mobius. 2004. Getting closer or drifting apart? *Quarterly Journal of Economics* 119, no. 3, 971-1009.

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

Star News Online. 2008. Associated Press picks top 10 stories of the year. (December 21).

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. 2010. Press coverage and political accountability. *Journal of*

Political Economy 188, no. 2, 355-408.

Stroud, Natalie Jomini. 2008. Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior* 30, no. 3, 341-366.

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.

Webster, James G. 2005. Beneath the veneer of fragmentation: Television audience polarization in a multichannel world. *Journal of Communication* 55, no. 2, 366-382.

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

A Appendix: Computing the Isolation Index

In this appendix we provide additional detail about the isolation index calculations that we present in the paper.

To compute the isolation index, we must classify all respondents as conservative or liberal. In the comScore PlanMetrix and MRI data, we classify those who answer "middle of the road" to the political outlook question as missing data and we classify all others as either conservative or liberal. In the GSS data, we classify moderates (point 4) on the 7-point ideology scale as having missing ideology data and we classify all others as either conservative (5 or more) or liberal (3 or less). In the CNES data, we classify moderates (points 5 or 6) on the 10-point ideology scale as having missing ideology data and we classify all others as either conservative (7 or more) or liberal (4 or less). As discussed in section 3.2, we assume that the share conservative among visitors to a given outlet with missing data is equal to the share conservative among those who declare a conservative or liberal ideology.

To estimate the isolation index for social interactions with acquaintances and political discussants, we define a separate "outlet" j corresponding to the acquaintances or discussants of every respondent i. We estimate the share conservative $\frac{cons_j}{visits_j}$ in each such outlet as the number of reported conservatives divided by the total number of reported conservatives and liberals. We then compute the sample analogue of equation (1) as the average of this share among conservative respondents minus the average of this share among liberal respondents, using the GSS and CNES sampling weights respectively.[6]

To estimate each individual's conservative exposure for the Internet using the 2008 comScore microdata, we define site j's share conservative to be the number of daily visitors who report conservative ideology divided by the number of daily visitors who report conservative or liberal ideology.

To estimate the isolation index for the Internet, offline media, and geographic areas, we define

[6]The sample of individuals we consider in the GSS and the CNES is the sample of respondents, rather than the sample of respondents' acquaintances / discussants. In the online appendix we report results that treat a respondent's acquaintances / discussants 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.

the sample analogue $\hat{cons}_j$ to be the number of observed conservative daily visitors to outlet j, divided by the share of all daily visitors to outlet j with non-missing ideology. We define $\hat{lib}_j$ analogously. We define the remaining sample analogues $\hat{cons}_m$, $\hat{lib}_m$, and $\hat{visits}_j$ of the terms in equation (1) by summing $\hat{cons}_j$ and $\hat{lib}_j$. We then compute the following estimate:

$$\hat{S}_m = \sum_{j \in J_m} \left(\frac{\hat{cons}_j}{\hat{cons}_m} \right) \left(\sum_{i \in I_{cons}} w_{ij} \frac{\hat{cons}_j - x_{ij}}{\hat{visits}_j - x_{ij}} \right) - \sum_{j \in J_m} \left(\frac{\hat{lib}_j}{\hat{lib}_m} \right) \left(\sum_{i \in I_{lib}} w_{ij} \frac{\hat{cons}_j}{\hat{visits}_j - x_{ij}} \right), \quad (2)$$

where x_{ij} represents respondent i's weight in estimating outlet j's share conservative, and $w_{ij} = \frac{x_{ij}}{\sum_{k \in I_{cons}} x_{kj}}$ for $i \in I_{cons}$ and $w_{ij} = \frac{x_{ij}}{\sum_{k \in I_{lib}} x_{kj}}$ for $i \in I_{lib}$.[7]

The terms $\frac{\hat{cons}_j - x_{ij}}{\hat{visits}_j - x_{ij}}$ and $\frac{\hat{cons}_j}{\hat{visits}_j - x_{ij}}$ are the share conservative among respondents *other than* i visiting site j, for the case where i is conservative and liberal respectively. We replace the share conservative $\frac{cons_j}{visits_j}$ in equation (1) with these "leave-out means" rather than with $\frac{\hat{cons}_j}{\hat{visits}_j}$ to avoid a small-sample bias discussed by Carrington and Troske (1997) and Ransom (2000). To see the intuition for the bias, note that the isolation index will be greater the more that $\frac{cons_j}{visits_j}$ varies across j. Even if $\frac{cons_j}{visits_j}$ is the same for all outlets (and hence isolation is zero), $\frac{\hat{cons}_j}{\hat{visits}_j}$ will tend to vary in a small sample, leading to an upward bias in the uncorrected estimator that uses $\frac{\hat{cons}_j}{\hat{visits}_j}$. Monte Carlo experiments confirm that the estimator in equation (2) is unbiased even when the number of sampled visitors per outlet is small, and that the uncorrected estimator has a clear positive bias. For reference, we present estimates of the uncorrected estimator in the online appendix.

B Appendix: Imputing the Ideology of Moderates

In our main calculations we assume that the share conservative among the moderates who visit a given outlet is equal to the share conservative among visitors to the same outlet who declare an ideology. In section 6.1.3 and the online appendix, we present results using a range of alternative assumptions.

In this appendix we investigate the plausibility of the assumption that we use in our main

[7] For Internet, we define x_{ij} to be constant across i and equal to $\left(\hat{cons}_j + \hat{lib}_j\right)$ divided by the number of Plan Metrix survey respondents with non-missing ideology who visit outlet j. For non-Internet media and geographic areas, we define x_{ij} to be i's MRI-defined sampling weight times the number of daily visits i made to outlet j, divided by the share of all daily visits to site j by respondents with non-missing ideology.

calculations, using data from the American National Election Study (ANES).

The ANES asks the following question of respondents: "We hear a lot of talk these days about liberals and conservatives. When it comes to politics, do you usually think of yourself as extremely liberal, liberal, slightly liberal, moderate or middle of the road, slightly conservative, extremely conservative, or haven't you thought much about this?"

In recent years of the study, respondents who report that they are "moderate or middle of the road" are asked "If you had to choose, would you consider yourself a liberal or a conservative?" About two-thirds of moderates declare a liberal or conservative leaning.

Appendix figure 1 shows the relationship, across US states, between the share conservative among moderates who report a leaning and the share conservative among those who declare an ideology initially. The assumption that the share conservative among a state's moderates is equal to the share conservative among those in the state who declare an ideology implies a slope of one in the fitted line presented in the figure. The estimated slope is 0.84, and is statistically distinguishable from zero ($p-value < 0.001$) but not from one ($p-value = 0.365$).

The evidence in appendix figure 1 shows that, if we can take moderates' expressed leanings as an indicator of their true ideology, the assumption we use is applicable for US states. We cannot test the assumption directly for Internet news outlets and the other media in our study because the ANES does not have detailed outlet-level visiting information comparable to the datasets we use in the paper.

Table 1: Size and Ideological Composition of Major News Media

US adult population: 42% conservative, 21% liberal, 38% moderate

Medium	Share of Daily Visitors Who Are:			Share of
	Conservative	Liberal	Moderate	Daily Visits
Cable	.45	.19	.36	.29
Local newspapers	.43	.19	.38	.29
Broadcast news	.42	.20	.38	.24
Internet	.37	.28	.35	.10
Magazines	.37	.28	.35	.05
National newspapers	.40	.31	.29	.03

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 of daily visitors who are [conservative/liberal/moderate] is the average across outlets of the share of daily visitors who report a given ideology, weighting each outlet in the medium by its average daily unique visitors. Conservative includes respondents who report that they are somewhat or very conservative; liberal includes respondents who report that they are somewhat or very liberal; moderate includes respondents who report that they are “middle of the road.” Internet data are from comScore; data on other media are from MRI.

Table 2: Size and Ideological Composition of Online News Outlets

Ten Largest

Site	Share of Daily Visitors Who Are:			Daily UV
	Conservative	Liberal	Moderate	('000)
drudgereport.com	.78	.06	.16	475
foxnews.com	.76	.10	.14	1,159
AOL News	.37	.23	.40	3,971
usatoday.com	.37	.25	.37	518
msnbc.com	.34	.26	.40	3,264
Yahoo! News	.31	.25	.43	6,455
cnn.com	.33	.27	.40	2,650
nytimes.com	.30	.45	.25	879
huffingtonpost.com	.22	.52	.26	583
BBC News	.16	.57	.26	472

Most Conservative

Site	Share of Daily Visitors Who Are:			Daily UV
	Conservative	Liberal	Moderate	('000)
billoreilly.com	.99	.00	.01	10
rushlimbaugh.com	.97	.01	.03	43
glennbeck.com	.89	.01	.09	38
humanevents.com	.91	.03	.06	21
townhall.com	.89	.04	.08	42
thestate.com	.58	.04	.38	36
aclj.org	.85	.06	.09	18
cnsnews.com	.92	.06	.01	12
drudgereport.com	.78	.06	.16	475
realclearpolitics.com	.87	.07	.06	41

Most Liberal

Site	Share of Daily Visitors Who Are:			Daily UV
	Conservative	Liberal	Moderate	('000)
thinkprogress.org	.05	.83	.12	12
blogcritics.org	.12	.61	.27	17
bvblackspin.com	.09	.43	.48	57
moveon.org	.14	.58	.28	21
BBC News	.16	.57	.26	472
blogtalkradio.com	.17	.58	.25	33
reddit.com	.15	.52	.33	36
newsvine.com	.21	.63	.16	56
alternet.org	.24	.67	.10	16
dailykos.com	.25	.68	.06	26

Notes: Average daily unique visitors is reported in 1000s. Data are from comScore. Conservative includes respondents who report that they are somewhat or very conservative; liberal includes respondents who report that they are somewhat or very liberal; moderate includes respondents who report that they are "middle of the road." "Most conservative" sites are those with the highest ratio of conservative to liberal daily visitors; "most liberal" sites are those with the highest ratio of liberal to conservative daily visitors. Sites are presented in descending order by the ratio of conservative to liberal daily visitors. 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 Ideological Composition of Offline News Outlets

Magazines

Magazine	Share of Daily Readers Who Are:			Market
	Conservative	Liberal	Moderate	Share
Barron's	.43	.19	.37	.02
U.S. News & World Report	.43	.20	.37	.14
BusinessWeek	.42	.21	.37	.07
Forbes	.40	.22	.37	.04
Fortune	.37	.24	.39	.03
TIME	.35	.27	.38	.31
Newsweek	.37	.29	.34	.27
The Economist	.35	.41	.23	.04
The Atlantic	.24	.55	.21	.01
The New Yorker	.17	.60	.24	.07

National Newspapers

Paper	Share of Daily Readers Who Are:			Market
	Conservative	Liberal	Moderate	Share
USA Today	.45	.22	.33	.40
The Wall Street Journal	.45	.21	.34	.29
The New York Times	.26	.54	.21	.31

Broadcast News

Channel	Share of Daily Viewers Who Are:			Market
	Conservative	Liberal	Moderate	Share
CBS	.42	.18	.40	.28
NBC	.44	.20	.36	.29
ABC	.42	.19	.40	.31
BBC	.37	.30	.33	.06
PBS	.32	.37	.30	.07

Cable

Channel	Share of Daily Viewers Who Are:			Market
	Conservative	Liberal	Moderate	Share
Fox News Channel	.54	.13	.33	.36
Bloomberg Television	.50	.18	.32	.01
CNBC	.41	.22	.37	.13
CNN	.40	.22	.38	.33
MSNBC	.39	.24	.36	.17

Data are from MRI. Conservative includes respondents who report that they are somewhat or very conservative; liberal includes respondents who report that they are somewhat or very liberal; moderate includes respondents who report that they are "middle of the road." Outlets are presented in descending order by the ratio of conservative to liberal daily readers/viewers. 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
Cable	.712	.679	.033
Magazines	.587	.540	.047
Local Newspapers	.695	.647	.048
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
Political Discussants	.796	.402	.394

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. Political discussants data are from the CNES. 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: Ideological Segregation by Site Size

Subset of Sites with Size Rank	Share of Daily Visitors	Cons. Exposure of Conservatives	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. Share of daily visitors is the fraction of total daily unique visitors across all sites accounted for by sites in the given size group. See section 3 for details on the construction of exposure and isolation measures.

Table 7: 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

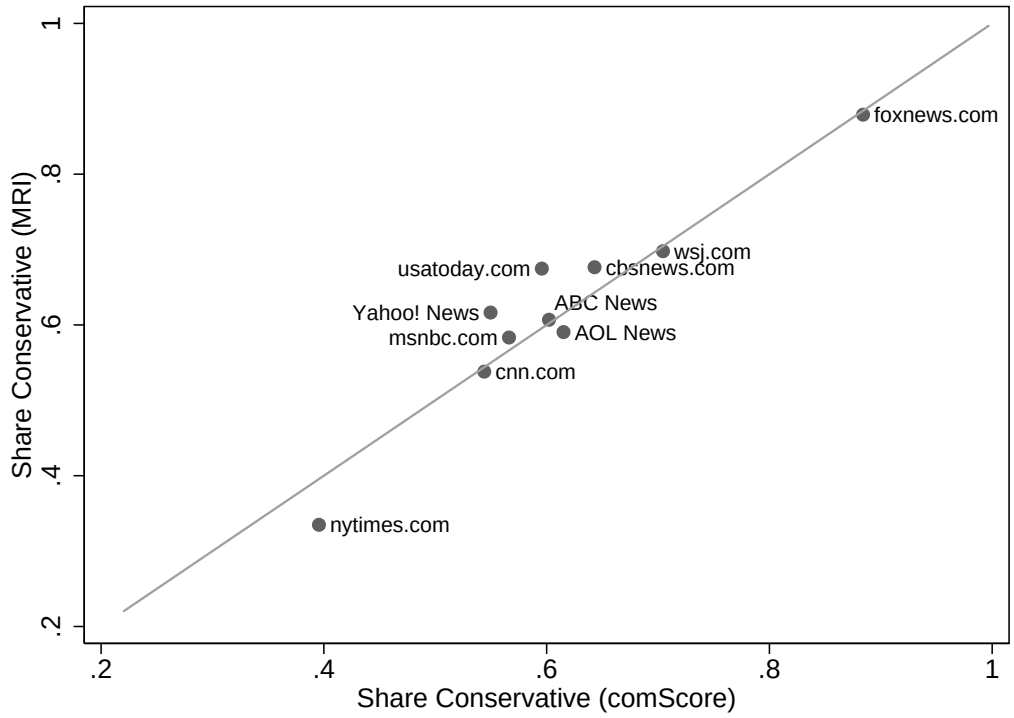
Notes: The table reports the share of all monthly unique visitors to a given site (listed in the first column) that make at least one visit in the same month to Yahoo! News, nytimes.com, and foxnews.com. These data are taken from comScore Media Metrix and are averaged over the 12 months of 2009.

Table 8: 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			
Add Google News	.601	.530	.072
391 Websites in comScore Microdata	.616	.517	.099
All News Websites Upper Bound	.617	.516	.102
Moderates			
Treat as Conservatives	.742	.692	.050
Treat as Liberals	.425	.346	.079
Drop	.618	.528	.091
Treat as Drawn at Random	.598	.551	.047
Treat as 50-50	.574	.525	.048
Alternative Ideology Measures			
Political Party	.522	.451	.071
Right-of-Median Zipcode	.510	.497	.013

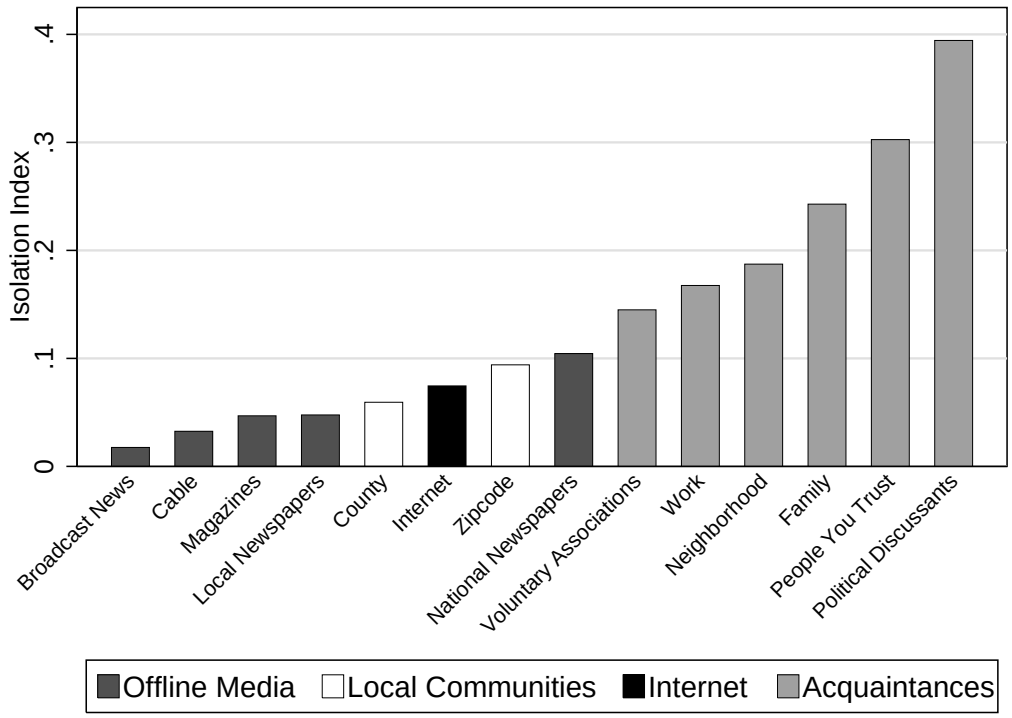
Notes: Data are from comScore. See section 3 for details on the construction of exposure and isolation measures. Zipcode ideology measure is constructed from Federal Election Commission data on political contributions. See section 6 for details.

Figure 1: Comparison of MRI and comScore Share Conservative



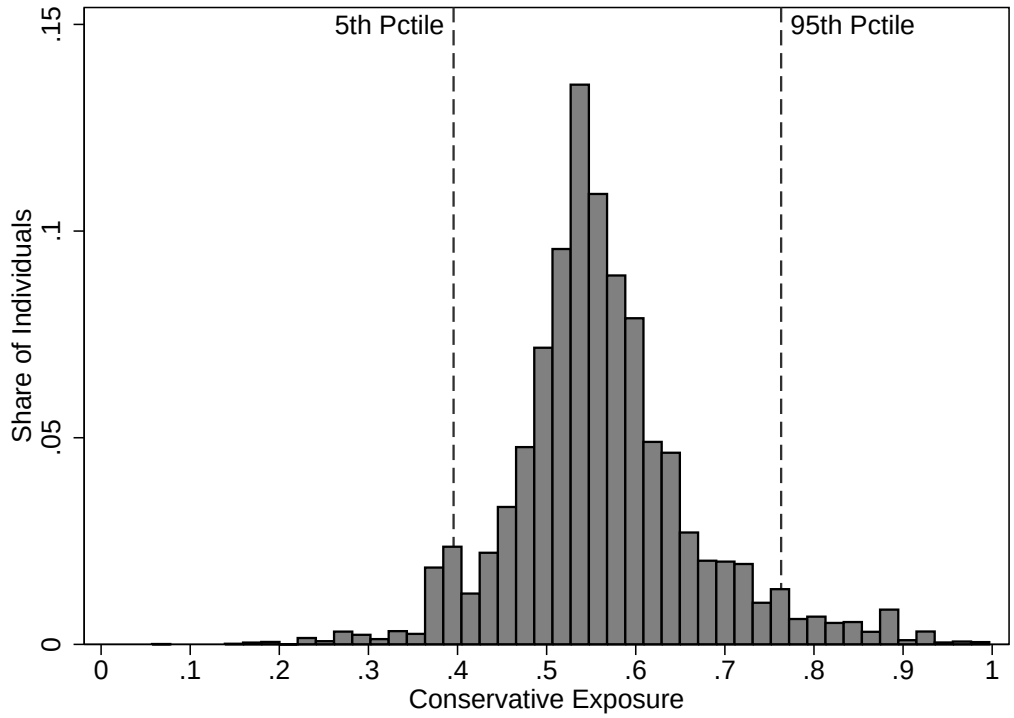
Notes: Data are from comScore and MRI. Share conservative is the estimated share of daily visitors who are conservative among those who report being either conservative or liberal. The line shown is the 45-degree line.

Figure 2: Ideological Segregation by Medium and Type of Interaction



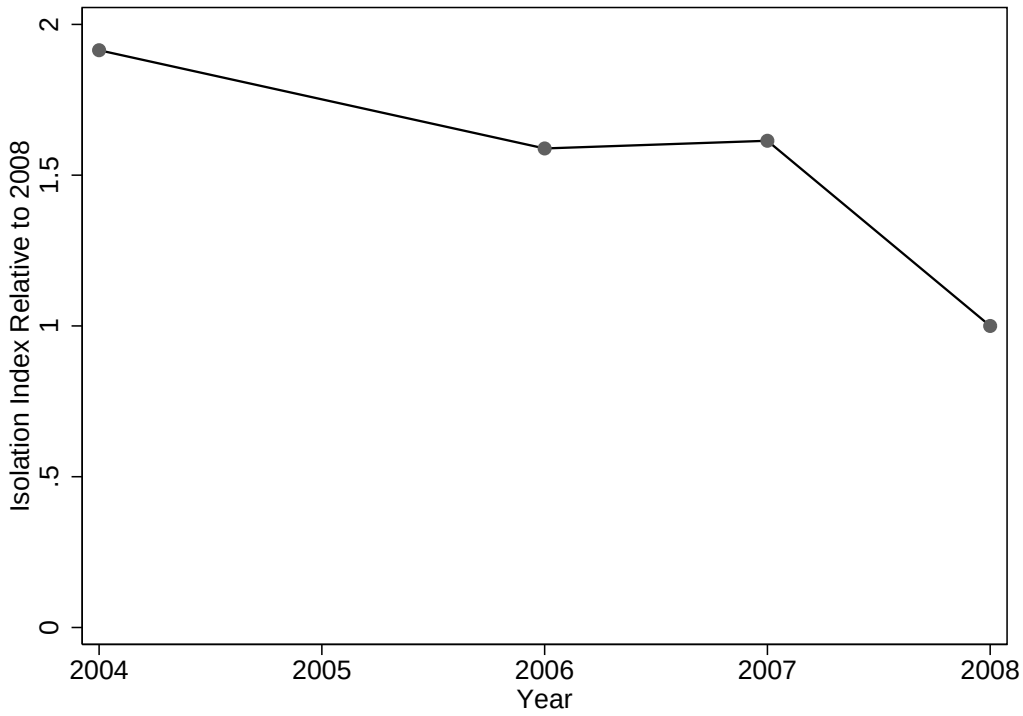
Notes: Internet data are from comScore. County, zipcode, and offline media data are from MRI. Voluntary associations, work, neighborhood, family, and "people you trust" data are from the GSS. Political discussants data are from the CNES. See section 3 for details on the construction of the isolation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

Figure 3: Distribution of Conservative Exposure across Internet Users



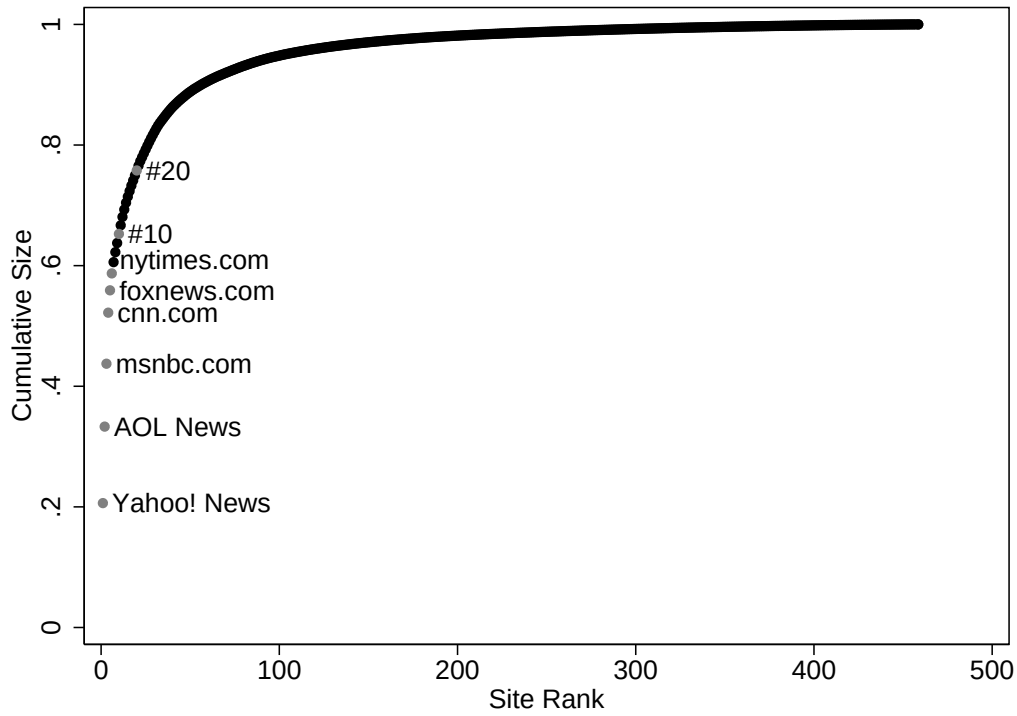
Notes: Data are from comScore. An individual's conservative exposure is defined as the average share conservative on sites she visited during 2008, weighting each site by the number of days in the year on which she made at least one visit to the site. See section 3 for further details on the construction of the exposure index.

Figure 4: Changes in Isolation Over Time



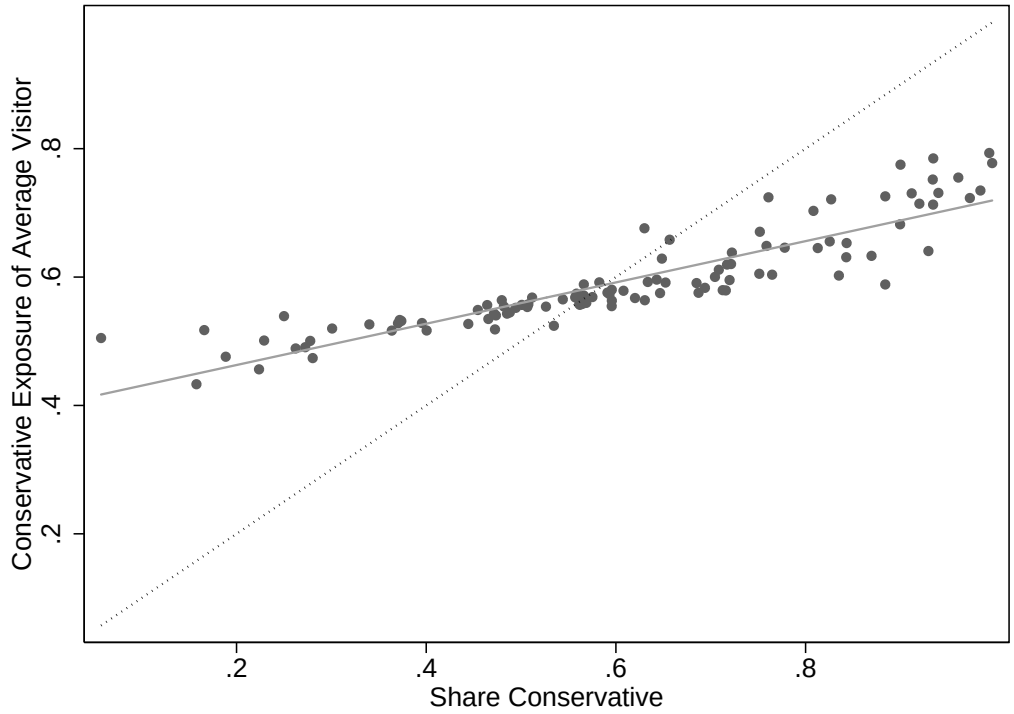
Notes: Data are from comScore microdata. The isolation index is scaled relative to the year 2008, so that the value for 2008 is 1.00.

Figure 5: Cumulative Distribution of Internet Unique Visits



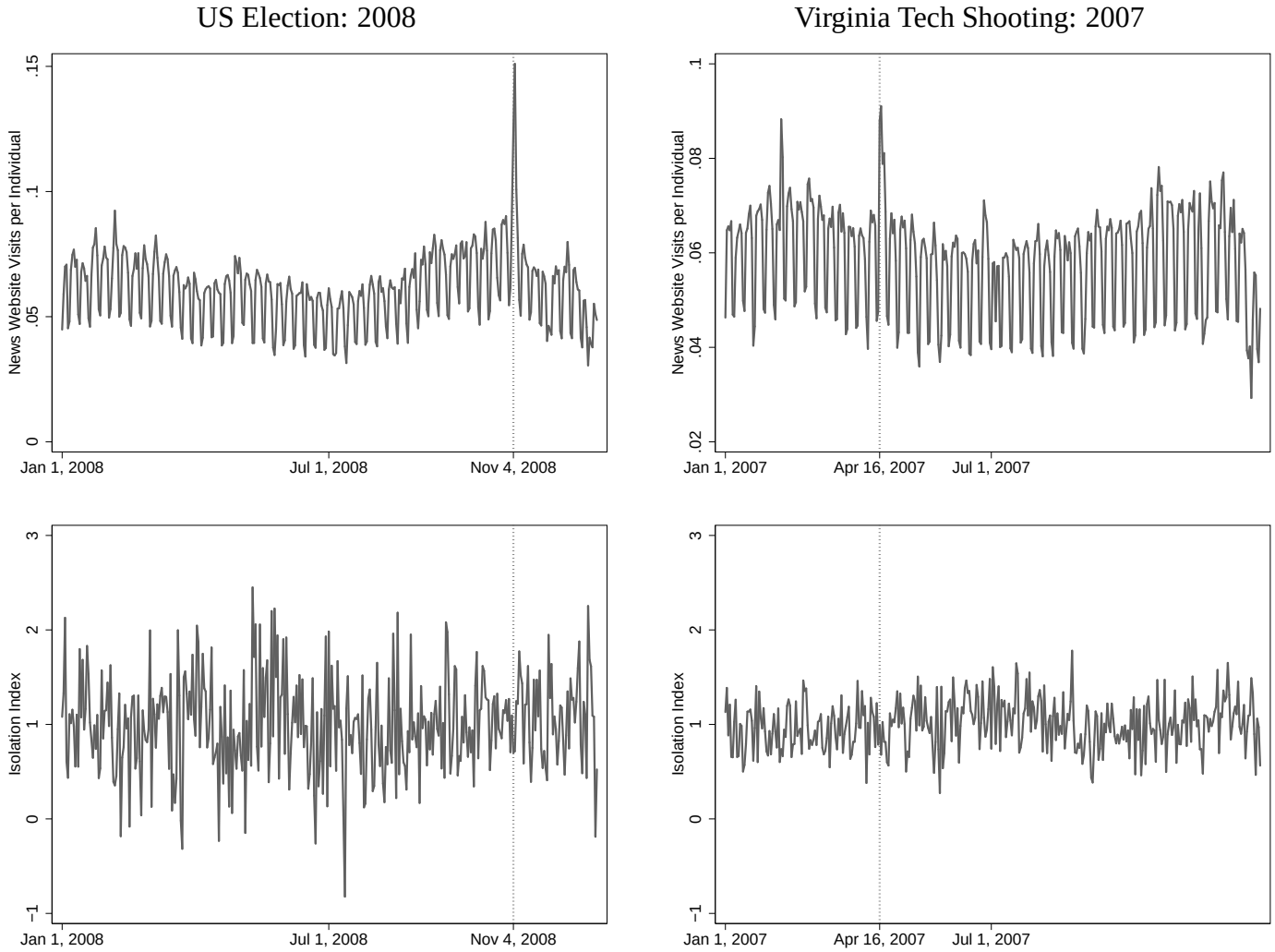
Notes: Data are from comScore. Size is measured by average daily unique visitors.

Figure 6: Visitor Exposure vs. Site Share Conservative



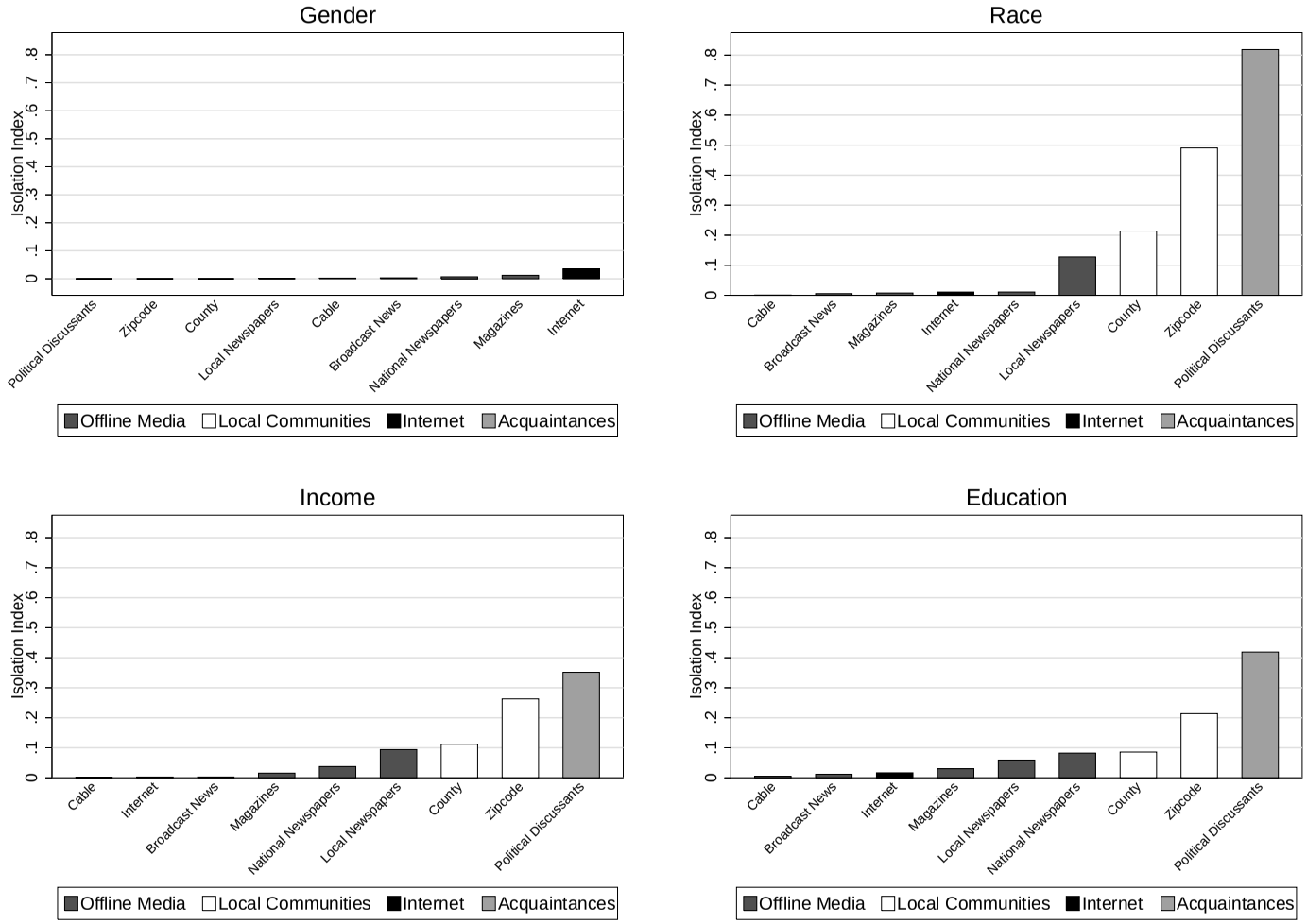
Notes: Data are from comScore. Figure plots conservative exposure of average daily visitor against the share of daily visitors who are conservative. An individual's conservative exposure is defined as the average share conservative on sites she visited during 2008, weighted by the number of days in the year on which she made at least one visit. The solid line is an OLS regression fit; the dotted line is the 45-degree line. See section 3 for further details on the construction of the exposure index.

Figure 7: Online Daily Visitors and Segregation by Day



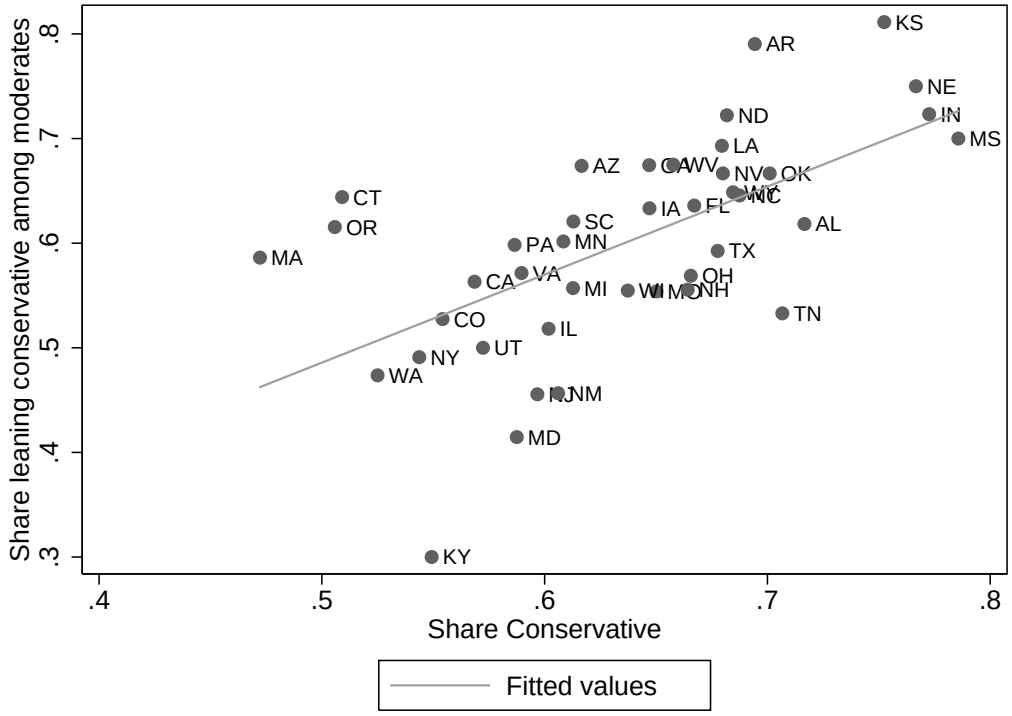
Notes: Data are from comScore microdata. In top panel, news website visits per individual is the average across individuals of the number of news websites in our main sample visited on each day. In bottom panel, the isolation index is scaled so that its mean across days is equal to 1.

Figure 8: Demographic Segregation by Medium and Type of Interaction



Notes: For each demographic characteristic we divide respondents into two groups to compute the isolation index defined in section 3. To compute gender isolation, we divide respondents into males and females according to the gender of the respondent in comScore PlanMetrix (Internet), CNES (Political Discussants) and MRI (geography and other media). To compute racial isolation, we divide respondents into blacks and whites (excluding all others) according to the race of the respondent in 2008 comScore microdata (Internet), CNES (Political Discussants) and MRI (geography and other media). To compute income isolation, we divide respondents into those with above- and below-median income according to the household income of the respondent in 2008 comScore microdata (Internet), CNES (Political Discussants) and MRI data (geography and other media). To compute education isolation, we divide respondents into those who graduated from a four-year college and those who did not according to the highest education level in the household in 2002 comScore microdata (Internet) and MRI data (geography other media), and the education of the respondent in CNES (Political Discussants). We truncate negative values of the isolation index at zero.

Appendix Figure 1: Imputing the Ideology of Moderates



Notes: Data are from the American National Election Study (www.electionstudies.org), years 1988, 1992, 1996, 1998, 2000, 2002, 2004, 2008. The unit of observation is the US state. The x-axis shows the fraction conservative among those declaring an ideology. The y-axis shows the fraction conservative among moderates who report a "leaning." Sample excludes states with fewer than 25 moderate "leaners" during the sample period.