IS THE INTERNET CORRELATED WITH POLARIZATION? A STUDY OF THE RELATIONSHIP BETWEEN INTERNET USAGE AND POLARIZATION

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Public Policy
in Public Policy

By

Chenlu Xu, B.S.

Washington, D.C.
April 9, 2020

IS THE INTERNET CORRELATED WITH POLITICAL POLARIZATION? A STUDY OF THE RELATIONSHIP BETWEEN INTERNET USAGE AND POLARIZATION

Chenlu Xu, B.S.

Thesis Advisor: Andrew Wise, Ph.D.

ABSTRACT

The internet as a communication medium benefits internet users by providing experiences difficult or impossible to encounter in real life. Nonetheless, it can also limit individual exposure to homogeneous political viewpoints and could be a primary reason for political polarization. Following the approach of Boxell et al. (2017), I examine the impact of internet use on different polarization indexes using the American National Election Survey datasets. I further substantiate the role of education and political interests in fostering political polarization and provide recommendations for mitigating political polarization.

TABLE OF CONTENTS

List of Figures

List of Tables

Introduction

Social media and other technologies promise a new source of information to internet users. In addition, by embedding algorithms in the software, technology companies are able to provide more advanced real-time communications and news to their users. As a result, communications can be tailored for each user as a result of predictions of which items users will most enjoy consuming. They also create a vision for users to connect with those who have similar preferences around the world. But this opportunity for connection also has a potential negative side. In 2011, Eli Pariser coined the term "Filter Bubbles" to describe the potential for online personalization to provide only liked-minded opinions, and therefore effectively to isolate people from a diversity of viewpoints. Similarly, T. Nguyen et al., (2014) argued that online recommendation systems tend to expose users to a narrowing information over time.

The present study builds on this work, examining how the internet traps U.S. users in separate bubbles, contributing in turn to political polarization. In recent decades, many studies have shown increased polarization among voters. For example, in a 2011 paper, Gentzkow explained polarization by using favorability toward parties; he found an increase in polarization in the U.S. beginning in the mid-1990s, exactly the moment when the internet became a mass phenomenon. Similarly, the annual split ticket voting survey conducted by the American National Election Survey (ANES) indicate that the proportion of voters voting for different parties in presidential and congressional elections decreased from 30 percent in 1972 to 15 percent in 2016.

In recent decades, as polarization has become salient, many authors have explored its potential linkage with social media. For example, B. Caplan (2008) found that voters tend to cast their votes mostly based on emotional factor and are influenced by the expressive value rather than the instrumental value of a policy. Moreover, today's internet users are prone to seek expressive value in social media and news, and to determine their voting behaviors based on their personalized information sources. Social media allow like-minded people to find one another and stoke each other's prejudices and grievances. Commenting on this relationship, former President Obama (2016) also said, "The capacity to disseminate misinformation, wild conspiracy theories, to paint the opposition in wildly negative light without any rebuttal – that has accelerated in ways that much more sharply polarize the electorate and make it very difficult to have a common conversation."

However, while polarization is widely deployed, the definition of polarization varies. Boxell, Shapiro and Gentzkow (2017) gathered and examined different measures of polarization from previous studies. They tagged different measures in ANES survey data and defended their arguments by analyzing polarization using various definitions. My study follows their method of constructing datasets and merging the indexes that reflect direct relationships with polarization. Using those indexes as response variables, I construct linear models to discover properties of polarization based on demographic variables in ANES survey.

In the next section, I present background and a review of relevant literature. Next, in the Theoretical Model section, which follows, I explain my model. After showing my results, I present my conclusions and policy recommendations.

Background

As the awareness of polarization rising, traditional theories explaining that news consumption does not lead to polarization have been challenged. Based on the theories of consumer preference, voters, who are also consumers in the policy market, have different preferences for different public goods and policies such as impeachment, or healthcare reform. Different individuals can disagree on the optimal level of a given policy because they benefit differently from it (Azzimonti and Fernandes, 2018). This is partly because they have different levels of income and wealth, which allows them to distribute their preferences fpr different policies in different ways. In other theories, voters are unaware of the optimal choice when they are presented with a bundle of policies. Examples can be environmental policy, restrictions on immigration, or choosing an optimal candidate in a presidential election. Since there are informational frictions, people's preferences may differ at the time their decisions are taken. But with the news media keeping people informed, the massive polarization shouldn't exist in the long term.

However, researchers have observed polarization that is intense and sustained. Various studies have provided measures of multiple aspects of polarization. In 1960, ~5% of Republicans and Democrats reported that they would "[feel] 'displeased' if their son or daughter married outside their political party"; by 2010, nearly 50% of Republicans and >30% of Democrats "felt somewhat or very unhappy at the prospect of interparty marriage" (Iyengar et al., 2012). The relative feelings of favorability of party members toward their own party increased by >50% between 1980 and 2015 (Gentzkow, 2016).

Ideological polarization has not only attracted the attention of traditional economists and political scientists, but also many researchers in the fields of psychology, computer sciences, and information technology. As today's polarization has strongly challenged the hypothesis of traditional theories, the rational ignorance theory (Downs, 1957) has seemed to me an increasingly creditable description of voter behavior. While voters do not gain explicit and direct benefit by absorbing news and political information, there is only a small chance that they will support a candidate on rational grounds, and that they will consequently benefit from the policy resulting from their decisive vote. Instead, voters choose the easy way, based on biased information which aligns most closely with their chosen beliefs. Zúñiga et al. (2017, p. 3) call this the "news-finds-me-perception." They argue that today individuals believe they can stay informed indirectly about public affairs – despite not actively following the news – through general internet use, information received from peers, and connections within online social networks. As a result, social media, which aimed to save time and improve efficiency for internet users in gaining information, have had a major impact on public opinion.

It's conventionally agreed that people's political views can be changed by creating an echo chamber, where beliefs are amplified or reinforced by communication and repetition inside a closed system. In a world of technology, we generally stay informed about the world around us by mediated text messages, online articles, and TV shows instead of direct personal experience. Hence, in recent times there has been a keenly contested debate in recent times that continues fiercely today around ideological polarization related to patterns in media and information consumption.

Literature Review

A. Internet Medium Approach

Long before addiction to social media became prevalent, researchers became worried about ideological segregation based on selective media. Even before Pariser (2011) argued that an informational dystopia was arriving as a result of filter bubbles, researchers were warning that search engines, news aggregators, and social networks were increasingly personalizing content through machine-learning models (Agichtein et al., 2006).

With the prevalence of filter bubble theories, more and more researchers began to study social media users, and to analyze their distribution across the political spectrum. Most of these researchers adopted the web Application Program Interface (API) to script datasets directly from Twitter and Facebook, which is an effective method for collecting enormous amounts of data. Some then used matching to investigate on the public available profile of user behavior in past elections. For example, Bakshy et al. (2015) used Facebook datasets to compare the ideological diversity of news and opinion shared on Facebook with that shared by individuals' friend networks. They also compared this ideological diversity with the sub-set of stories that appeared in individuals' algorithmically ranked News Feeds.

While distinct Web links (URLs) are the majority of the user communications that Bakshy and his colleagues observe, they distinguished stories as either "hard" content (such as national news, politics, or world affairs) or "soft" content (such as sports, entertainment, or travel) by training a support vector machine on unigram, bigram, and trigram text features. By averaging the ideological affiliation of each user who shared on article, they were able to observe substantial polarization among user-shared hard content, with the most frequently shared links clearly aligned with these users' political ideologies.

To reflect the effect of cross-cutting content on different political views, Bakshy et al. also looked at the affiliation of users' networks. While individuals do not generally encounter information at random in offline environments (Sears and Freedman, 1967), or on the internet (Flaxman et al., 2013), they found that liberals tend to be connected to fewer friends who share information from the other end of the ideological spectrum. This is opposed to Mason's (2015) finding, that liberals tend to share more cross-cutting content and have more friends from the opposing party. Bakshy and his colleagues identified three stages in the media exposure process by investigating the effects of content that is (i) potentially available on the network, (ii) truly exposed to users, and (iii) finally selected by users. They found that in the context of the Facebook, individual choices limit exposure to attitude-challenging content more often than algorithms. This finding provides a basis for my own analysis in this paper.

Twitter is another widely discussed social medium. However, giving that tweets are limited to 140 characters, researchers have cast doubt on the ability of analysis of such communication to tease out real voter sentiment or intent (Bermingham and Smeaton, 2011). Nonetheless, while many researchers agree that evidence of an echo chamber effect should be more explicit on Twitter, Twitter's open access to APIs provides an excellent environment for information diversity research (An et al., 2012). For example, Bozdag et al. (2014) and Barberá et al. (2015) conducted their studies using public data collected from Twitter. To determine whether online communication is more than an echo chamber, Baberá and his colleagues analyzed hashtags from previous political and non-political events. They passively collected data on users' follows, retweets, and social networks, an approach which, in their view, avoids bias based on behavior exhibited in laboratory experiments. By using a latent space model of political ideology, they assume that the probability of a connection between two individuals, both nodes of a given network, is negatively related to their distance in a latent ideological space.

In addition, both studies used data from users' following political accounts to project the estimated political position for the specific user. Bozdag et al. (2015) developed political news accounts in the Netherland and Turkey, and Barberá et al. (2015) added accounts of Obama for America, The Huffington Post, Ready for Hillary, Stephen Colbert and others on the liberal side, as well as Rush Limbaugh, Bobby Jindal, Tucker Carlson, and the Tea Party, which are popular among conservatives. Bozdag et al. observed that 73% of the left side users retweeted only from left users and replied to left users, while 72% of the right side users did the same. The situation was more extreme for Turkey, where 93% of left side users only retweeted from and replied to left side users, and 94% of the right side users showed the same behavior.

In addition to political views and events, the two studies added non-political events to determine whether trends in retweets told the same story as political events. They analyzed tweets on the 2014 Super Bowl, 2014 Oscars, Winter Olympic Games, and other events based on the same political spectrum, and concluded that online communication structures are flexible and situation-specific, and that the aggregate level of political polarization depends heavily on the nature of the issue. In terms of individual idea exchange among tweeters with highly similar ideological preferences, the authors found significant traces. When it came to explicitly political issues, individuals were clearly more likely to pass on information that they received from ideologically similar sources than to pass on information that they received from dissimilar sources. Also, while there was no observed polarization on social and entertainment events, some specific events exhibited a dynamic shift from national conversation to echo chamber – such as in the case of the tragic elementary school shooting in Newtown, Connecticut, when the conversation shifted from the tragedy to a debate over gun control policy. The researchers found similar patterns in tweets related to the possibility of military intervention in Syria. This finding bolsters my construction of the *internet use* variables to represent political news consumption.

In another approach to assessing the echo chamber on the internet, Flaxman et al. (2016) examined web-browsing histories. Using the Bing Toolbar, a popular add-on application for the Internet Explorer web browser, they collected data on user behavior via an opt-in agreement. Their findings, based on 50,000 users, showed that users that regularly read partisan articles are almost exclusively exposed to only one side of the political spectrum. In this sense, nearly all users exist in so-called echo chambers.

With more advanced data analytics, other authors identified detailed fluctuations in echo chamber effects. They also criticized the use of surveys on the grounds that self-reported survey responses are subject to measurement error and social-desirability bias (Gentzkow and Shapiro, 2011). However, Barberá (2015), later on, validated his survey-based results by pointing out that these results were highly correlated with statewide averages of ideology calculated on the basis of survey and socio-demographic data (Lax and Phillips, 2012).

B. Survey Questionnaires Approach

The limitations for using Facebook and Twitter datasets are obvious. The most recent survey of news consumption by the Pew Research Center found that 62% of adults in the US get their news on social media, and 66% of all Facebook users use the platform for news consumption (Gottfried and Shearer, 2016). However, while not all people use Facebook and Twitter, whether Facebook and Twitter user groups are representative of the general public is unknown. Studies show that, compared with the U.S. population as a whole, Facebook's users tend to be younger, more educated, and more often female (Maeve et al., 2013). Moreover, while Boyd and Crawford (2011) have stated that, "many journalists and researchers refer to 'people' and 'Twitter users' as synonymous", some users have multiple accounts, some accounts are shared by several people, and some people never create an account and access Twitter via their internet browser. In addition, retweets and followers do not always represent user ideology. For example, some users retweet or reply with bad intentions, such as trolling.

In addition, despite the views of authors cited in previous paragraphs, systematic bias won't be canceled out by the large number effect. To overcome this analytic problem, Boxell et al. (2016) used more generalized observations from yearly surveys conducted by the American National Election Survey and Pew Research Center from 1948 to 2016 to predict the impact of the internet. Although the number of respondents each year was limited because of resources, Boxell et al.'s results were more convincing than those of studies based solely on social media data because they were plausibly representative of national polarization. Boxell and his colleagues collected measures of polarization from different journals and created nine measures corresponding to variables in the surveys. Averaging the nine measures, they created an index to detect traces of polarization. To their surprise, they found that young age groups with a higher proportion of social media users had lower polarization indexes. Their results also showed the selection bias of internet-collected data: people selected themselves into the social media user sample.

Other critics of internet-collected data focus on whether internet use truly provides a narrower viewpoint to users. For example, Barberá (2015) claimed that social media reduce mass political polarization by its “weak ties”. He referred to the connections that one usually would not normally encounter offline that can be exposed by algorithms. To some extent, he argued social media expand the homophilic nature of personal networks. In addition, he pointed out that most online activities examined in empirical studies only represent a small proportion of the time citizens spend online, which is now increasingly devoted to visiting social media sites such as Twitter or Facebook, particularly among young adults.

Similarly, Mutz (2002) argued that individuals on the internet are increasingly exposed to a diversity of opinions, the experiences should lead to social consensus and higher political tolerance. He used Twitter data from countries where a large proportion of citizens use Twitter (e.g., Germany, Spain, and the United States) in order to diminish the effect of non-representative users. He also linked Twitter profiles in the United States to publicly available voter files through name identification. This enriched his datasets with a variety of measures of offline behavior, and successfully supported his assumption that social media provide a variety of political viewpoints rather than an echo chamber.

C. *Human Behavior or Machine Algorithm?*

One can argue that what Mutz did not necessarily eliminate sampling bias since there are always people who do not use the internet, and so that can never be included in the sample pool.

In addition, whether a broader viewpoint across all topics on the internet can be introduced to politics is questionable. That may be the main reason Mutz's findings are inconsistent with the findings of other authors using internet sampling as their approach. Among these articles mentioned above, only a few (e.g., Bakshy, 2015) realize that personalized algorithms function based on user interactions, which means that, if users want to trap themselves in an echo chamber, it is naïve to denigrate algorithms. Most authors have passively accepted the idea that the formation of filter bubbles creates an echo chamber. Few have commented on the reason for that formation, which is user behavior. The addiction to liked-minded opinions is the main reason that algorithms are able to create cocoons, which trap people inside.

Sophr (2015) pointed out that selective exposure online can be a source of ideological polarization. Selective exposure refers to the fact that individuals have a tendency to consume media which align with their views and beliefs and which avoid content that is different from their perspective or challenges their positions (Frey, 1986; Stroud, 2008). Sophr also discussed the psychological cause of selective exposure – confirmation bias. The term refers to seeking or interpreting evidence that is partial to existing beliefs, expectations, or a hypothesis in hand (Nickerson, 1998, p.175). While individuals may not come online only to confirm their beliefs, the influence of confirmation bias can help algorithms to create echo chambers. Iyengar and Hahn (2009, p.21) pointed out that some studies may fail to distinguish between deliberate or motivated exposure and "de facto" exposure that was a by-product of voters' personal networks or social context. Hence, it is hard to reject the conclusion that social media algorithms function like traditional media. For example, competition forces newspapers to cater to the prejudices of their readers, and greater competition typically results in more aggressive catering to such prejudices as competitors strive to capture market share (Mullainathan and Rubin, 2008, p.106). As a result, the power of supply and demand reinforces the cycle through which social media construct a personalized filter bubble for every consumer using these media. To mitigate this effect, citizens and scholars have proposed redesigning the algorithms. Some developers are also looking for solutions to reduce the likelihood of echo chambers; for example, with tools that anonymize users' online actions so they cannot be tracked by customization software (Difranzo and Gloria-Garcia, 2017).

Experiments have also showed that people are not subjectively bound within their filter bubbles, and that the same algorithm can have different influences on different people. Thus, in 2016, The Guardian newspaper conducted an experiment in which they asked 10 U.S. voters to log in to accounts opposing their political viewpoints during the final month leading up to the presidential election. While some individuals confirmed their pre-existing views about the other side, others found greater understanding from the experiment and stated that they needed to be more proactive about finding diverse news (Wong et al., 2016). Similarly, Feezell (2018) recruited a randomly selected list of 5,000 registered students from the University of New Mexico, a large public university. These students were offered periodic political information and then asked about their level of issue salience. The author found that the effect of this information was strongest among those with low levels of political interest, who were most likely to avoid political information in favor of more entertaining options, such as social media.

D. Present Study

Recognizing that the impact of social media differs among different people, my paper adds to the literature by studying demographic variables that are not emphasized in previous studies. Because online web-scripted datasets do not include demographic status, it is impossible to exclude the biases of representative study groups. On the contrary, the ANES survey questionnaires provided sample weights, which reduces or eliminates concern about sampling bias. In contrast to Boxell et al.'s work, I therefore primarily look into whether demographic groups are represented among respondents, and analyze whether patterns vary based on education level, gender (Maeve et al., 2013) and race. On the one hand, my results can help to prove or disprove conclusions of the literature that analyzes patterns of online accounts. On another hand, my findings on education level and attention to politics can help to determine whether there are differences in selective exposure effectiveness, based on the assumption that more education can mitigate the impact of selective exposure. Though more educated citizens may still be affected by filter bubbles, their polarization index should be much lower.

Descriptive Data

My data come from the American National Election Studies (ANES). My primary sources are the ANES 1948-2012 Time Series Cumulative, 2008 Time Series Study, 2012 Time Series Study, and 2016 Time Series Study datasets. ANES surveys are face-to-face and online surveys of the voting-age population that conducted in both pre-election and post-election phases. They contain numerous demographic variables and political measures. To examine the relationship between social media and elections, I restricted the 1948-2012 cumulative file to presidential election years, and merged pre-election and post-election data.

Since 2012, there has been a separate sample of respondents who completed ANES survey data online. To make my sample groups more representative, I also include those respondents. As I noted, I used weights from published survey files to conduct a non-biased estimate for the entire population.

I conducted the data wrangling using Stata software. ANES provided major commands to recode the measurement variables. I deleted from my dataset original responses recorded as “not applicable” and “don’t know”. Then I recoded numerical variables so that they make sense for analysis. For instance, I recoded a 7-point scale from 0 to 6 representing “strongly agree with democrats” and “strongly agree with republicans” as -3 to +3. For the measures I seek, I also deleted missing observations. My measures are selected from Boxell et al.’s nine measures. Since my purpose was not to show a universal trend toward polarization. I did not follow their method of averaging all measures to create an index, but instead selected distinctive measures that can explain polarization in different ways. The final dataset contains more than 3,000 valid and measurable observations.

Descriptive statistics for all major variables are included in table 1 below.

Table 1. Descriptive Statistics

Variable name	Variable	Count	Mean	Sd	Min	Max
respondent_id	ANES recorded respondent ID	19324	33053.05	22250.43	1	55674
year	Election year	19324	2008.871	6.127603	1996	2016
male	Respondent gender	9771	0.195067	0.3962726	0	1
age	Respondent age	17002	48.61587	17.21114	18	97
race	Respondent race	16897	1.907617	1.54985	1	6
south_state	Whether R is in a south state	17002	1.657452	0.4745757	1	2
education	R's education level	16885	3.914125	1.216483	1	6
iapLib	Ideology affect polarization -- liberal	14642	51.71807	24.45309	0	100
iapCon	Ideology affect polarization -- conservative	14715	56.5141	22.90146	0	100
iapD	Ideology affect polarization -- Democratic	16707	55.4464	28.26719	0	100
iapR	Ideology affect polarization -- Republican	16665	45.89337	26.99469	0	100
elect_interest	R's election interest	15876	2.163832	0.7454163	1	3
perceivedD	Perceived ideology -- Democrat	15801	-1.121891	1.528821	-3	3
perceivedR	Perceived ideology -- Republican	15720	1.220611	1.598136	-3	3
pres_house_split	Split vote in 2016	9178	0.1380475	0.3449686	0	1
has_internet	whether R has internet access	7566	0.6286016	0.4832106	0	1
internet	Whether R view political information online	8248	0.4283463	0.5472301	-5	1
extreme	R's opinion on how extreme R is	12985	0.1933	1.504379	-3	3
healthIns	R's opinion on health insurance	10016	-0.1174121	1.960059	-3	3
econJobs	R's opinion on economic and job market	13661	0.2521777	1.830045	-3	3
blackAid	R's opinion on black aid	14235	0.6454514	1.837535	-3	3
govServe	R's opinion on government services	13034	0.0299985	1.659435	-3	3
defSpend	R's opinion on defence spending	13040	0.2754601	1.558004	-3	3
selfIdeo	R's self reported ideology	12985	0.1933	1.504379	-3	3
selfParty	R's self reported party afliiation	16885	-0.363814	2.108889	-3	3
issue1	R's rating on specific political issue 1	12853	3.752431	2.284098	0	9
issue2	R's rating on specific political issue 2	9772	0.2025174	1.471324	-3	3
issue3	R's rating on specific political issue 3	9407	2.073669	2.3818	0	9
poliKnow	R's political knowledge	4722	2.228864	1.191567	1	5
weight1	Cumulative weight	17002	0.9661643	0.7437771	0	6.4445
weight2	Face to face interview weight	12853	1.000169	0.7469748	0.0212	4.9909
internet_pres_campaign	New format question: R view online politics	2322	0.8561585	2.620961	-2	5
opinion_count	R's count of opinions	14235	5.006533	1.267574	1	6
finaliap	Ideology affect polarization	14491	27.56062	27.38343	0	100
final_perceived_party	Perceived party ideology	15658	3.160685	1.66164	0	6
final_political_opinion	Political opinions	6537	1.37456	0.6373842	0	3
final_vote	Split ticket voting	9178	0.8619525	0.3449686	0	1
final_partisan	Partisan sorting	12937	-0.3759759	13.23196	-97	14

Figure 1 presents the weighted average numbers for each dependent variable across the election years. All polarization measures increased in this duration. Since the ANES survey collects its sample from the entire population, the weighted averages should represent the polarization trends for both internet user groups and the non-internet user groups.

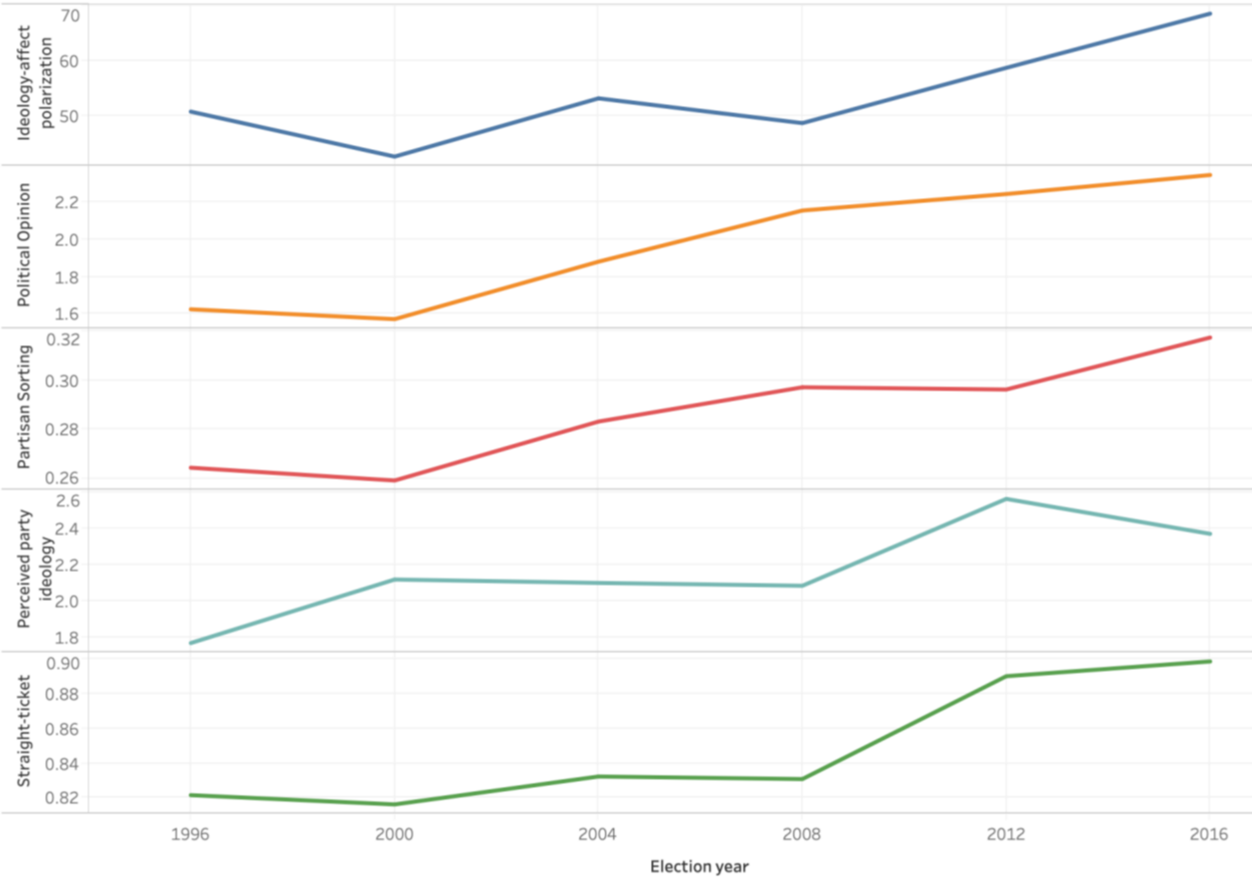

Figure 1. Dependent – Polarization Measures

Theoretical Model

To examine the relationship between the use of Internet use and polarization, I created the following model:

$$Polarization = f(Internet, Demographic\ Characters, Individual\ Characters, Time, \mu) \quad (1)$$

I added factors of time and demographic cohorts to represent the change in different groups. The individual characteristics represent behavioral factors affecting use of the internet and social media. I selected these factors because of their potential relationship with ideological effects. Internet use, demographics, individual characteristics, and polarization are factors that contain multiple variables. When it comes to internet use and demographic characteristics, I intend to look for general trends in the convergence of increasing polarization factors in time, and the divergence in social-demographic features.

My goal in constructing this model is not to explain the causes of polarization. In fact, the causation factors are too complex to be examined in linear regressions. Despite the impact of social media, researchers agree that many other factors that are hard to measure can also lead to a polarized political atmosphere. For instance, polarization can increase as a result of gerrymandering or the manipulation of electoral borders to favor a political party. Party and public polarization can also be a self-reinforcing process. And hence, whether the party, the public, or social media is the leading indicator is unknown. Therefore, it is difficult to identify a causal relationship using linear regressions. Given this problem, my model seeks different trends among different characteristics in both demographic and individual inclinations. And those trends represent only an explicit relationship, not causation. Whether demographic factors are significant will help to explain the importance of studies using API methods. It will also help to identify divergence among different age, race, and gender groups in previous works on polarization. In addition, individual factors will help in evaluating whether people's choices affect differences in polarization. As a result, the analysis will help to explain studies focusing on psychological rationales for polarization, and also studies trying to discern the relationships between machine algorithms and human behavior.

Empirical Framework

A. Polarization Factors

Although researchers agree on what polarization represents, the number of methods that can measure polarization quantitatively is extremely large. In the API articles I mentioned in the literature review, most authors developed an index based on data that they can collect, but the equations of their measurers vary. I examined the measures of polarization developed by Boxell (2018) and selected the most widely used five polarization measures. I reconstructed these measures precisely and only intentionally deviate where explicitly stated. As stated in the section above, party polarization can also affect public opinion. Thus, I mainly used polarization measures that measures ideology affiliations.

"Ideological affect polarization", a measure developed by Iyengar et al. (2012) and Gentzkow (2016), is constructed based on ANES thermometer ratings of parties and ideologies to capture respondents' feelings toward those on the other side of the political spectrum. The measure is the sum of the mean differences, taken separately for liberals and conservatives, between the favorability of individuals toward their ideological group and their favorability toward the opposite ideological group. This measure is presented in equation (2).

$$\frac{1}{weight_i}\sqrt{[weight_i\left(index_i^{liberal} - index_i^{conservative}\right)]^2} \tag{2}$$

"Partisan sorting", a measure developed by Mason (2015) and Davis and Dunaway (2016), captures the association between self-reported partisan identity and self-reported ideology. This measure is defined as the average absolute difference between partisan identity and ideology (both measured on a 7-point scale), after weighting by the strength of partisan and ideological affiliation. This measure is presented in equation (3) and (4).

$$\frac{weight_i}{105}[g(|party_i - ideology_i| + 1)(|party_i + 1|)(|ideology_i + 1|) - 7] \qquad (3)$$

$$g(x) = max_{i \in \cup_t}(\lceil party - ideology \rceil + 1) + min_{i \in \cup_t}(|party - ideology| + 1) - x \qquad (4)$$

"Perceived partisan-ideology polarization" captures the extent to which respondents perceive ideological differences between the parties. The measure is defined as the average perceived ideology of the Republican Party minus the average perceived ideology of the Democratic Party, each on a 7-point liberal-to-conservative scale, where $\tilde{R}_i$ and $\widetilde{D}_i$ denote on individual's perception of how conservative the Republican or Democratic party is, according to the following equation:

$$weight_i(\tilde{R}_i - \widetilde{D}_i) \qquad (5)$$

"Political opinions" measures the extent to which individuals' issue positions line up on a single ideological dimension. This measure is the average absolute value of the sum of six responses, with each valid response defined as conservative, moderate, or liberal. The three are coded as -1, 0 and 1. The responses are based on six survey questions including: aid to blacks, foreign defense spending, the government's role in guaranteeing jobs and income, government health insurance, government services and spending, and abortion legislation. Here, k denotes a response to one of the six politics questions.

$$weight \left| \sum_{k} k_i \right| \qquad (6)$$

"Straight-ticket voting" captures whether individuals split their votes across parties in an election. The measure is defined as the survey-weighted proportion of voting respondents who report voting for the same party in both the presidential and House elections of a given year.

B. Internet Use

Internet use is a factor that potentially suffers from endogeneity in my model. As I discussed above, researchers don't generally agree on the factors that caused polarization. While both sides of the arguments seem plausible to me, it is difficult to claim that my polarization factors do not exhibit a reverse causal link with my independent variable. I therefore used a more objective description, which is household internet access to instrument on my independent variable. Using this instrument has several advantages. First, internet access is strongly correlated with viewing political information online, and this correlation weighs more than the correlation between internet access and other dependent polarization measurements. (See Table 2.) This supports the exclusion restriction for using instrumental variables.

Also, the internet has become a pervasive good that now is less correlated with social and economic status than it was when it first appeared. Therefore, it has become a better instrument compared to when ANES first conducted its survey. However, I am also aware of a potential problem with using *Internet access* as my instrumental variable. It can be correlated with education level, which is one of the major controls in my model. It is arguable that more educated people have a greater possibility of having internet access. However, the Pearson R-correlation between education and internet use is 0.092, which is trivial compared to the relationship between *Internet access* and my independent *Internet* variable.

Table 2. Validation of Instrument

	Internet access
Internet access	1
View political information online	.571
Ideological affect polarization	.04
Perceived party ideology	.069
Political opinion	-.001
Straight ticket voting	.015
Partisan sorting	.103

C. Formal Regression

My formal regression is estimated in two stages:

First stage:

$$Internet = \rho_0 + \rho_1 gender + \rho_2 race + \rho_3 year + \rho_4 education\ level + \rho_5 political\ knowledge + \rho_6 Internet\ access + \varepsilon \qquad (7)$$

Second stage:

$$Polarization^* = \alpha_0 + \alpha_1 Internet + \alpha_2 gender + \alpha_3 race + \alpha_4 year + \alpha_5 education\ level + \alpha_6 political\ knowledge + \mu \qquad (8)$$

Here I expand the full regression model, with polarization factors remaining as a general indicator. The reason is measures of polarization cannot be run together in individual regressions. And *Polarization* is marked with an asterisk to denote that it is composed by several measures. I also expand demographic and individual factors to survey-question-level variables, with each variable representing one question from the survey directly. The year variable measures the time of each election year from 1996 to 2016. Education level indicates a respondent's education experience at the time they were interviewed. Based on the ANES survey questionnaire, respondents have been divided into unevenly distributed bins such as "high school diploma", "some college", and "college degree."

My political knowledge variable comes from the survey question set testing a respondent's knowledge about politics. This variable indicates whether the respondent is attentive to political information, and can further support the work of Feezell (2018.) In the years 1996 to 2012, those five questions are conducted using these questions: whether respondent correctly named the majority party in the House of Representatives before the election, whether respondent correctly named the majority party in the house after the election, whether respondent remembered his house representative's name, and whether the respondent remembered one of his Senators' names. In 2016, the variable is based on a question for interviewers: how well they think their respondent understood political issues on a 5-point scale. Similarly, the *Political knowledge* during 1996 to 2012 has been recoded to a 5-point scale.

The reason for my *Political knowledge* variable is to express a respondent's interest in reading political information. If a respondent scored a zero over the years, he or she might be more interested in online information other than politics. One can argue that choosing specific questions such as how well a respondent understands abortion and healthcare policies would be more effective. But the distribution of that variable cluster mostly around zero, which means that most people do not understand those specific policies in detail. In addition, my approach also helps to differentiate *Political knowledge* from *Education level*.

I expect my internet use variable to be positive and significant in all regressions since previous studies cited in paper generally agree that internet use is positively correlated with polarization. Boxell et al. measured only face-to-face interview records to maintain a consistency in those measures. They also measured a period beginning in 1948; so online and phone responses do not account for a large proportion of their dataset. However, since my instrumental variable requires the prevalence of internet as an assumption, I only focus on the years from 1996 to 2016, and the proportion of respondents that did the survey via phone or the internet become salient for my study. More data points from the phone interview and online interview responses should make my correlation linkage between internet user and my polarization indexes more representative than Boxell et al.'s (2018) relationships. Boxell et al.'s measures of polarization show different but generally consistent trends. I expect to encounter similar differences in my regressions since each measure I select in this paper is designed to reflect a distinctive aspect of polarization. Nonetheless, some polarization indexes may become less salient when online and phone respondents are merged into the dataset.

My demographic controls should indicate divergence among different gender, race and education level groups. Since polarization is a universal phenomenon, my assumption on internet use should have indifferent effects for each internet user, regardless of their gender or race. However, it is worth noting that polarization, may be more likely to appear in female users' filter bubbles. Therefore, my *Gender* variable may show non-significant effects in 1996, but become negative in later years when filter bubbles become more affluent. I expect my *Education Level* to be significant and negative, since previous studies mentioned that more education will help to diffuse the effect of the echo chamber. My *Political knowledge* variable should be strongly positive since more interest in learning about politics makes individuals more vulnerable to polarizing online content.

Results

Three of my five comprehensive regressions show significant relationships between my polarization indices and use of the internet. Regressions using *Ideological-affect polarization*, *Perceived party ideology*, and *Straight ticket voting* measurements result in positive, large, and statistically significant coefficients on internet use. The other two regressions show some non-significant results with explainable signs. The results on my control variables in the five regressions together tell a consistent story. Table 3, column (1) presents regression results using *Ideological-affect polarization* as the polarization measurement. While the treatment variable is instrumented on whether respondents have internet access, I observe a positive 6.989 coefficient estimate on internet use. In other words, based on a 100-point scale of a thermometer rating of ideologies, respondents who frequently view political campaign and information online are 6.989 points more polarized than respondents who don't view online information. This finding is significant at the 99% confidence level. This result supports my hypothesis. Moreover, given the fact that ideology measurement is most frequently discussed in the related literature, the results also show that using ideology to explain polarization may be the most effective method to evaluate polarization.

Table 3, column (2) indicates a non-significant effect of 1.795 points for the *Partisan sorting* measurement, an index built by Mason (2015) and finalized by Davis and Dunaway (2016). Mason used two weak measures of social identity to construct this measurement: individual self-reported party strength, and ideology strength. His measure of partisan sorting, which is built to capture both the alignment of respondent partisan identity and the strength of identity, has not been thoroughly explored. Since the effects of sorting on bias and anger are stronger than the effects of sorting on activism, people could become more biased and angrier than they are actually polarized. This may explain the insignificant results observed in this regression, because people who demonstrated their biased speech online could act as cynics or not vote. This echoes the concern noted in my literature review that online platforms are not suitable for evaluating polarization because of fake identities, internet bots, and sharable accounts.

Table 3. Main Regression Results

	(1) Ideological affect polarization	(2) Partisan sorting	(3) Perceived party ideology	(4) Political opinions	(5) Straight ticket voting
View political information online	6.989***	1.795	0.757***	-0.103	0.234*
	(2.238)	(1.580)	(0.153)	(0.107)	(0.142)
Male	1.102	-0.492	-0.061	-0.053	0.097
	(1.032)	(0.652)	(0.066)	(0.043)	(0.147)
Education	1.543***	0.479*	0.176***	-0.025	-0.077
	(0.421)	(0.268)	(0.028)	(0.017)	(0.049)
Election year	0.018	0.013	-0.005	0.018**	0.030***
	(0.084)	(0.054)	(0.005)	(0.008)	(0.010)
Political knowledge of respondent	3.501***	-0.826***	0.187***	0.011	0.132***
	(0.336)	(0.243)	(0.023)	(0.018)	(0.048)
Internet access					-0.100
					(0.144)
Constant	-31.710	-25.707	11.547	-34.401**	-57.955***
	(168.215)	(108.643)	(10.900)	(16.300)	(19.551)
Obs.	4724	3306	4552	1780	3005
Pseudo R_2	.z	.z	.z	.z	0.013

Standard errors are in parenthesis
**** p<0.01, ** p<0.05, * p<0.1*

In contrast, as shown in column (3), the simplified version of Partisan sorting, which is *Perceived partisan-ideology polarization*, shows a significant correlation with internet use. This measurement consists of a mathematical expression: the strength of the two parties' ideologies on a 6-point scale. Respondents are asked to place marks on this scale indicating how left, or right they think the two parties truly are. I observe a positive coefficient of 0.757, with a standard error of 0.153, significant at the 99% level. This suggests that respondents who view political information online experience an ideological strength of parties that is significantly different from those who do not. Compared to the latter, they think the two parties are 0.757 points on average farther away from each other. This result is consistent with the findings of my first regression, where the result of internet use is also significant at 0.01 alpha level.

Like the *Ideological-affect polarization* index, *Perceived partisan-ideology polarization* also measures the extent of ideological distance. In addition, using self-reported rates on both sides, it captures respondents' evaluation of party ideological alignment. While in my *Ideological-affect polarization* regression I only take individuals' ratings of their opposing sides into account, *Perceived partisan-ideology* polarization measures the favorability of individuals towards their own partisan group and the opposed group, which is differentiated from the opponent evaluation method used by *Ideological-affect polarization*. The identical results of both regressions further prove the impact of the internet on polarization.

Table 3, column (4) reports the results of my regression using the *Political opinion* index as my measure of polarization. The results show a negative but statistically non-significant relationship with internet use. Although the sign is different from my hypothesis, the standard error of this estimate is large enough for us to understand the story behind a measure that folded in five issue opinions. many authors argue that issue position polarization is not at all synonymous with social polarization. It is possible that we observe polarization in each issue area independently, while the whole story cannot be inferred based on one or two issues. The measurement of Political opinion increases its validity by using not one, but five issue opinions. And the five selected issues are among the most well-known points of controversy that have been discussed in the political arena since 1972, which is the year of the first records for these five questions in the ANES survey. However, many features of this measurement remain unclear , including whether these five issues always reflect a trend toward polarization or if any one of the five issues explain arguments other than polarization. For instance, the arguments for aid to blacks represent an issue with different stakeholders compared to the topic of this thesis. With different values and standards figuring into healthcare policy, the definition of healthcare policies can also change and evolve. And therefore, I believe I observed an insignificant estimate based on that part of the story so that internet usage impact is not significantly different from zero.

While the two significant regressions in Table 3 may only explain ideology and not activism, my fifth regression, a logit regression, measures whether respondents split their votes for the House and the President in election years. Here, the previous instrumental variable has become one of the controls. For my main coefficient of interest, I observed a 0.234 point estimate with 0.142 standard error, which supports the significance of internet usage on split-ticket voting. In other words, the log of the odds ratio between the internet user group and the non-internet user group is 0.234. The ratio of the odds for the internet user group to the non-internet user group is 1.26, which means that people who view political information online are 26% less likely to split votes. Although the low R-square may produce some doubt about the efficacy of this model, and some statistical noise remains based on the survey I used, this result is consistent with my hypothesis.

Table 4 presents the results of four regressions in specific years, testing the individual year-effect of internet use in 1996 and 2016 in order to trace the evolution of internet impact. In these regressions, I selected the two measures that provided significant outcomes above as dependent variables. The results are surprisingly consistent with the development of social media technology. While in 1996 social media algorithms were missing, the internet only explains a 13.049 point increase in polarization at the 90% level of significance, by 2016 this estimate had risen to 22.602 with a 99% level of significance.

Table 4. Specific Year Regression

	(1) Ideological affect polarization in 2016	(2) Ideological affect polarization in 1996	(3) Perceived party ideology in 2016	(4) Perceived party ideology in 1996
View political information online	22.602***	13.049*	1.477***	1.070**
	(5.555)	(6.963)	(0.474)	(0.452)
Male	2.343	0.074	-0.031	0.132
	(2.229)	(1.876)	(0.130)	(0.129)
Education	1.253	2.255***	0.223***	0.135***
	(1.252)	(0.540)	(0.076)	(0.043)
Political knowledge of respondent	3.035***	3.509***	0.146**	0.282***
	(1.128)	(0.546)	(0.062)	(0.046)
Constant	-3.163	1.242	0.833***	1.215***
	(5.412)	(2.223)	(0.315)	(0.185)
Obs.	981	1433	950	1383
R-squared	-0.098	0.062	-0.005	0.064
year	2016	1996	2016	1996

Standard errors are in parenthesis
**** p<0.01, ** p<0.05, * p<0.1*

The same story can be found in my regressions on *Perceived party-ideology* polarization. Whereas in 1996, I can only observe a 1.07 point increase in polarization, at the 90 level of significance, by 2016 I discover a 1.477 point increase at the 99% level of significance. In all of these regressions while the time effect has been filtered out, the impact of internet use has substantially increased compared to the comprehensive model I discussed above. This suggests that in the absence of the effects of time, the impact of internet on polarization is much more than we see in real life. However, it is unwise to suggest that time diluted the effect of internet use. Regressions on specific years only suggest a static view of polarization and may suffer from omitted variable bias.

Evidence supporting this result can also be found in the estimates of my control variables. In regressions (1) and (3), in which I measured the impact of internet use with statistical significance, year controls are not significant. In the regression on *Political opinions*, where the results are insignificant, year cohorts become the only statistically significant variables that explain the increase of the polarization index over time. Year cohorts in other regressions mostly show a positive impact. My logit model on straight ticket voting also represents a year effect that's significant at the 99% level, suggesting that during election years, people are 3% more likely to vote straight tickets than split tickets. This result improves the credibility of my hypothesis that the increase in polarization in recent years is significantly correlated with internet use.

While the year cohort explains polarization over time, my other control variables shed light on other polarization arguments. Conventional views suggest education as a way to become more informed about social policies and politics, and hence, become less polarized. However, the results in all my significant regressions on ideological-affect polarization, partisan sorting, and perceived party ideology, education is correlated with increased polarization. All three statistically significant coefficients suggest that well-educated people are significantly more polarized than the others. While the education variable is not evenly divided in years of education, but instead in terms of the level of education, I think one possible explanation is that this comes from the different gaps between each categorical level of education. While “some college without a degree” may not indicate a big difference from “college degree”, respondents with a “high school diploma” may never decide to receive higher education. Another possibility is that more educated people tend to participate in online campaigns and political discussions more often than people who do not require electronic devices as a necessity for their daily jobs. As a result, online information may have more effect on those who are more educated.

Another meaningful variable is political knowledge. As explained in the discussion of my empirical models, political knowledge affects whether a person wants to learn more about politics. Hence, I suspected that it is related to polarization. My results show that knowledgeable voters are more polarized. And three of my five regressions show a positive impact of political knowledge on polarization, which contradicts my expectation. The effects differ based on the scale of indices, but they are all significant at the 99% level. In my *Partisan sorting* and *Political opinion* regressions in which I failed to control with a precise definition of polarization, political knowledge is either not significant or shown to have a negative impact. To explain this, I point to the randomness of unknowledgeable respondents. Responses on polarization of those who know less about the political spheres should be a random composition of numbers centered on zero. In contrast, more knowledgeable respondents could present a polarization that is becoming radical.

Furthermore, people with more political knowledge are more likely to look for online content and so they are likely to enjoy more content delivered by personalization algorithms.

Discussion and Policy Implication

A. Study Limitations

The evidence presented in this study generally agrees with the view that the internet is one of the significant causes of polarization. Some shortcomings of the study have been mentioned above. Here I summarize three key points that can be inferred from these shortcomings and provide suggestions on how to improve and make useful implications to this field. First, internet use can be a major cause of polarization, but the effect of the printing press, television programs and radio stations are not accounted for in this study, and hence my models suffer from various omitted variable biases that could have has a larger impact compared to internet use. The point estimates on the internet can be diverted from the influence from other sources I brought up.

In another way, the effect of the internet can also be mixed up with other polarization sources. For example, major media and television all have their public accounts on social media, and they have been posting op-eds and political arguments since they realized the capacity of online markets. What's more, the impact of internet use has been changing with technology development. While examples like filter bubbles are some causes of polarization, they do not represent the sole reason why polarization seems to be common throughout the internet. Specific year regressions also suggested that while in 1996 we have no filter bubbles and algorithms, respondents were less influenced by internet use. With the evolution of smartphones and the emergence of social media as intermediaries connecting with the political spectrum, the influence of the internet has become more pervasive. However, back in 1996, while there were no complete algorithms like Facebook and Twitter use today, the internet was still inducing polarization. And if the internet already had become influential on polarization when it only let people see what they wanted to click and read, shouldn't we assume that people's natural intention to explore contents related to themselves was already causing polarization, and therefore, the internet is only a catalyzer? This corresponds to previous arguments made by Caplan (2008) that people tend to become careless of politics when it exceeds their emotional capacity for acceptance. In other words, these people are anti-intellectuals who only favor easily interpreted content that most excites their nerve systems. In fact, the measurement of *Political opinion* first created by Abramowitz and Saunders (2008) is based on the five most simply understood questions on the whole political spectrum, and therefore only summarized the partial causation of internet impact.

Second, the method I used is based on survey respondents. Their self-reported values may introduce a degree of social desirability bias, and people tend to perform differently once they know they are selected to a panel. I have discussed the solution for this issue, relying on web-scripted data, which would eliminate these study biases, but would also suffer from omitted variables that cannot be observed through the internet. Previous studies have noted that incomplete data have been their biggest problem, and in some research designs, nearly half of their datasets are assumed biased on average or on other respondent properties.

In addition, polarization observed from the internet may be worse than it is in real life. After all, internet users can behave very differently compared to what they might do in actual life. Rowland (2011) suggested that a person's understanding of the political world, as well as the vehemence with which he or she reacts to political events, can be drastically different from how he or she really acts in voting and social life.

While survey data and web-scripted data seem like the best of both worlds that scientist never achieve, I suggest using randomized control trials as an experimental method to conduct the research. An experiment targeted to solve the myth of polarization by the internet will solve the lack of comprehensive datasets for good, and can provide more implications on other potential influences on polarization. Experiments that are similar to those of Wong et al (2016) and Feezell (2018) but taking a larger scope of observation in numbers and diversity of respondents would be ideal.

As stated above, people can observe polarization on the internet and think that two sides are starting a civil war tomorrow, but in real life, the issue is never that severe. The methods to evaluate polarization vary, much like different people's view on political issues vary. And as a result, the evidence presented in the section above only makes sense when it comes to polarization indices on left-right ideology. That is my study's third limitation, which is that if omitted variables are real problems, the results can be biased. Also, if my two significant regressions on Ideological-affect polarization and Perceived party-ideology polarization do not represent the true story of polarization, regressions for individual years can only lead to further misconceptions.

B. Major Implications of Polarization Studies

Based on peer studies mentioned above, my results may be useful to be interpreted in this study field and policymakers. First, I have confirmed that polarization is not a pervasive phenomenon that affects each individual in the United States. The increasing trend of polarization can be mainly explained by the internet, instead of race or gender. This finding disagrees with studies mentioned in my literature review section using data from the internet to argue that polarization is everywhere. In addition, the trends of increasing polarization based on different measures also disprove the view that polarization is only a myth. My results show that polarization is a widespread issue grounded on internet use, regardless of users' gender or ethnicity. Gottfried and Shearer (2016) found that over 62 percent of adults in America consume news on social media. That doesn't necessarily prove that polarization only exists among 62 percent of Americans. Rather, polarization is more of an online problem than one that is only cared about by internet trolls. Because the democratic system involves each stakeholder in this issue, the real number of the population using the internet affects polarization more than internet user groups. Other people who are not even internet users will also likely be affected throughout their social network channels.

Other features of polarization presented above also underline specific reasons polarized groups are generally more educated and informed. As I noted in Sophr's (2015) argument above, people face selective exposure because they select content for themselves. Technological algorithms are only tools that utilize the weakness of humanity: people enjoy seeing what they like and want to avoid been contradicted. A way to understand this is to treat every internet user as a consumer in an economic market. While consumers in the goods-market tend to maximize their utility with limited monetary resources, internet users also maximize their utility with the limited time they have. While entertaining content is already consumed at a high rate compared to political information, people are less likely to treat tweets, Facebook posts, and feeds that cost them less than three minutes to read as a part of the self-critical approach. Nonetheless, this sense of informal communication conveyed by Twitter is different from reading scholarly work and trying to learn from complicated jargon and complicated logic to improve internet users themselves.

In my results, the significance of political knowledge bolsters Sophr's argument that psychologically selective exposure is one of the major causes of polarization. Since people observe their identities as superordinate based on this easily leaned internet use tendency, I argued a political knowledge variable is necessary for differentiating internet users interested in entertainment than those who do spend time on politics. Respondents who have a basic understanding of politics know what information they will enjoy. Those people are clear about what their preferences are in the internet information market, and therefore their future political acts will align with these materials they read. These acts such as split-ticket voting are effectively captured using the measures I have used.

In previous sections, I have noted Caplan's (2008) book discussing what generates voter participation in politics. Caplan argued that, with little ability to change the outcome of an election, individual voters participate in voting and spend time and energy to learn about politics mainly for emotional satisfaction. As we use political background knowledge as an index to express how much respondents are interested in consuming online political information, the inference becomes clear: Those who enjoy consuming online political information more will receive more emotional satisfaction from reading news coming out of filter bubbles, and thus become more polarized.

Did we blame technology too soon? I think it is uncertain. As I mentioned before, my model more focused on correlation and than on direct causal linkage. Indeed, it is as hard to prove consumer rationality caused polarization as proving filter bubbles caused polarization.

Those respondents at the time having a strong political knowledge and interests, how do we know whether they raised their interests from polarization itself? Perhaps the individual was an independent before a polarized article changed his mind. This remains unknown to me since it is not reflected in my dataset. However, what I can say with confidence is once this process happens, the internet becomes the main causation of self-reinforcement.

C. Related Stakeholders and Policy Issues

Four major stakeholders in the polarization issue are policymakers, media, technology companies, and internet user groups. Policymakers, as representatives elected by their constituents, should be responsive to these constituents. In the meantime, they also play a role in framing regulations on technology companies. Policymakers may also be affected by online content and expect their constituents are more polarized than they truly are. Internet user groups, on the contrary, contribute to social media to share their thoughts and vote for their representatives. Technology companies are financially driven, and hence seek to benefit their owners. Hence, they only refine these algorithms to attract users to put more time and interest into their products. The mass media also seek to attract audiences by creating content.

What are the features of internet user groups? I want to address the arguments based on Fiorina's (2006) idea that polarization in America is mostly a myth concocted by social scientists and media commentators. Fiorina's arguments condemn the parties and candidates who hang out on the extremes, and media coverage reporting what is polarized, for helping to create this myth. My findings also suggest that polarization is more generally discussed among elite groups with higher education and political knowledge than average. Other people not willing to spend time and stamina to learn about politics may be evenly spread out in the left-right axis. But with the catalyzing effect of the internet, the more they read, the more they are informed about one side, and the less they can be convinced by the opponent's opinions, regardless of whether their opponents' value align with their interests. The more individuals enjoy the emotional value gained from that content, the less they will care about the effectiveness and benefits of particular policies. This self-reinforcing process inadvertently crowds out valuable policies and reforms. This process that make careless constituents neglect and misunderstand valuable policies also affect their elected representatives. Boven (2018) noted the vicious cycle between policymakers and their constituents. His study shows that even though Republicans agree with ideas of climate-change, they still oppose policy solutions proposed by Democrats, since they hope to be aligned with their constituents and their party.

As mentioned above, polarization can be an elite problem since education is positively related to internet use. However, additional important features have not been fully explored using this dataset. For example, I found that although the education factor promotes polarization on average, the high end of the education variable tells a different story. While the mean effect of everyone with less than a master's degree is a positive 10.2 on *Ideology affect polarization*, the effect from internet use for masters and doctors is only 2.3. College degrees and high school degrees, overall, receive the highest impact from the *internet use* variable. This finding then counters Fiorina's argument that polarization is only an elite problem. Indeed, most influences from education seem to support this conclusion since the number of high school and college degrees are substantial compared to masters and above. Therefore, although the most educated people are less impacted, the entire internet user group still experiences a negative impact from education.

Identified as part of the private sector, the other two stakeholders, mass media and tech companies, claim to have no responsibility since they operate in a competitive free market. What they are doing in creating algorithms and writing editorials is essentially improving company profits as every private firm would do. However, there are examples of regulating media and technology companies in order to achieve less extremity in the free market. The European Union introduced the General Data Protection Regulation (GDPR) in 2018, and as a result, social media platforms face fines if they do not delete extremist content within an hour. YouTube has a self-governance policy to remove inappropriate content and demonetize videos with strong political opinions.

Does regulation of media and technology companies solve the problem for good? Based on my findings on polarization, regulation of social media does not necessarily deal with the problem of voter behavior. And therefore it becomes less cost-effective. In the following paragraph, I discuss a few options that may work better to address polarization.

When the emotional and polarized language used to galvanize one side directly antagonizes the other, intergroup contact can be a practical solution. This suggests that knowing out-group individuals can help to reduce prejudice between groups. One promising civic model for enabling more meaningful contact between groups is “Citizens Assemblies.” In 2017, the public policy school in University College London gathered 50 randomly selected individuals of the UK electorate in Manchester over two weekends. They made sure that participants represented the electorate in terms of age, gender, ethnicity, social class, place of residence, and voting in the 2016 Brexit referendum. Participants first learned about Brexit issues from experts and each other, then deliberated on the options before reaching a conclusion. During the assembly, researchers observed mitigations of misunderstandings, and the proportion who supported Brexit declined from over 50% to less than 50%.

Nonetheless, while hearing from what the internet keeps you from would be a potential solution that works, but it may not be practical for everyone who is not selected to attend a citizens' assembly. Therefore, improving voters' understanding of how specific policies would benefit them will help to set up a balanced discussion environment for them. For example, we can use data science to extract and summarize information from campaign finance records and other large datasets in order to provide better information cues to voters. In the meantime, providing more thoughtful, balanced, and respectful consideration of different viewpoints would encourage individuals to learn more about the out-group perspectives.

There have also been some suggestions for amending the democratic system. Berman (2016) recommended authorizing impartial commissions to draw election district boundaries in order to stop gerrymandering. Other suggestions such as adopting proportional representation, increasing turnout by lottery or other incentives, and reforming the campaign committees to strengthen their relationship with outside groups are all feasible ideas that could work based on my discussion of voter behavior. In other words, as long as the approach doesn't violate voters' incentive to participate in politics, the solution would theoretically work to reduce polarization between the two parties and their constituents.

D. Implications for Future Studies

Because of the limitation of current methodologies, no conducted experiment can testify the accurate causal inference about the effect of internet use, and specifically, the embedding algorithms in social media. In many studies, including mine, these two terms are treated as one because we cannot distinguish the effect of the two from available datasets. However, it might be feasible to conduct an experiment, which does not need a big sample size, but need enough time to let researchers decompose human behavior on internet use.

Second, many authors that have discussed polarization disagree on how people should view polarization. It was a problem a decade ago when arguments focused mainly on whether polarization was a myth. Now those indexes are emerging, but I think a more generally accepted, measurable definition can help to improve effectiveness in studying polarization significantly. The average value method used by Boxell is close to my expectations, but I still found some indexes difficult to interpret.

Third, the ANES dataset lacks specific regional identifiers to help me locate average polarization among states. A fixed-effect method would significantly improve my model in terms of homogeneities across different states. There are unobservable, unmeasurable features. For example, respondents from Texas may be more polarized than those who come from Virginia. By measuring the fluctuation in each state, researchers would be able to deal more effectively with omitted variable bias. In addition, during 20-year epoch of the ANES surveys, some respondents dropped out, and new respondents were added. It would be beneficial if ANES could adopt matching to gain some consistency between dropped out participants and new participants.

References

Agichtein, E., Brill, E., & Dumais, S. (2006). Improving Web Search Ranking by Incorporating User Behavior Information. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 19–26. https://doi.org/10.1145/1148170.1148177

Aldrich, J. H., & McGraw, K. M. (2011). *Improving Public Opinion Surveys: Interdisciplinary Innovation and the American National Election Studies*. Princeton University Press.

An, J., Cha, M., Gummadi, K., Crowcroft, J., & Quercia, D. (2012). Visualizing Media Bias through Twitter. *Sixth International AAAI Conference on Weblogs and Social Media*. Presented at the Sixth International AAAI Conference on Weblogs and Social Media. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4775

Anspach, N. M. (2017). The New Personal Influence: How Our Facebook Friends Influence the News We Read. *Political Communication*, *34*(4), 590–606. https://doi.org/10.1080/10584609.2017.1316329

Azzimonti, M., & Fernandes, M. (2018). *Social Media Networks, Fake News, and Polarization*. (Availability Note: Information provided in collaboration with the RePEc Project: http://repec.org).

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Barberá, P. (n.d.). *How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S.* 46.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, *26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Bermingham, A., & Smeaton, A. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results. *Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP 2011)*, 2–10. Retrieved from https://www.aclweb.org/anthology/W11-3702

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). *Is the Internet Causing Political Polarization? Evidence from Demographics*. (Availability Note: Information provided in collaboration with the RePEc Project: http://repec.org).

Bozdag, E., Gao, Q., Houben, G.-J., & Warnier, M. (n.d.). Does offline political segregation affect the filter bubble? An empirical analysis of information diversity for Dutch and Turkish Twitter users | Elsevier Enhanced Reader. https://doi.org/10.1016/j.chb.2014.05.028

Carpini, M. X. D., & Keeter, S. (1996). What Americans Know about Politics. In *What Americans Know about Politics and Why It Matters* (pp. 62–104). Retrieved from http://www.jstor.org/stable/j.ctt1cc2kv1.6

Davis, N. T., & Dunaway, J. L. (2016). Party Polarization, Media Choice, and Mass Partisan-Ideological Sorting. *Public Opinion Quarterly*, *80*(S1), 272–297. https://doi.org/10.1093/poq/nfw002

Degroot, M. H. (1974). Reaching a Consensus. *Journal of the American Statistical Association*, *69*(345), 118–121. https://doi.org/10.1080/01621459.1974.10480137

DiFranzo, D., & Gloria-Garcia, K. (2017). Filter bubbles and fake news. *XRDS: Crossroads, The ACM Magazine for Students*, *23*(3), 32–35. https://doi.org/10.1145/3055153

Elkin, F. (1957). Personal Influence: The Part Played by People in the Flow of Mass Communications. By Elihu Katz and Paul F. Lazarsfeld. With a Foreword by Elmo Roper. A Report of the Bureau of Applied Social Research, Columbia University. Glencoe, Ill.: The Free Press. 1955. Pp. xxii, 400. $6.00. *Canadian Journal of Economics and Political Science/Revue Canadienne de Economiques et Science Politique*, *23*(4), 572–574. https://doi.org/10.2307/139030

Feezell, J. T. (2018). Agenda Setting through Social Media: The Importance of Incidental News Exposure and Social Filtering in the Digital Era. *Political Research Quarterly; Salt Lake City*, *71*(2), 482–494. http://dx.doi.org/10.1177/1065912917744895

Flaxman, S., Goel, S., & Rao, J. M. (2013). *Ideological Segregation and the Effects of Social Media on News Consumption*. https://doi.org/10.2139/ssrn.2363701

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, *80*(S1), 298–320. https://doi.org/10.1093/poq/nfw006

Frey, D. (1986). Recent Research on Selective Exposure to Information. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 41–80). https://doi.org/10.1016/S0065-2601(08)60212-9

GENTZHOW_WP1_polarization-in-2016.pdf. (n.d.). Retrieved from http://www.techpolicy.com/TechnologyAcademicsPolicy/media/Miscellaneous-images_1/GENTZHOW_WP1_polarization-in-2016.pdf

Iyengar, S., & Hahn, K. S. (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Iyengar, S., Sood, G., & Lelkes, Y. (2012). “Affect, not ideology: A social identity perspective on polarization”: Erratum. *Public Opinion Quarterly*, *76*(4), 819–819.

Lax, J. R., & Phillips, J. H. (2012). The Democratic Deficit in the States. *American Journal of Political Science*, *56*(1), 148–166. https://doi.org/10.1111/j.1540-5907.2011.00537.x

Maeve, D., Aaron, S., & Inquiries, D. 20036USA202-419-4300 | M.-857-8562 | F.-419-4372 | M. (2013, December 30). Social Media Update 2013. Retrieved November 10, 2019, from Pew Research Center: Internet, Science & Tech website: https://www.pewresearch.org/internet/2013/12/30/social-media-update-2013/

Mason, L. (2015). “I Disrespectfully Agree”: The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science*, *59*(1), 128–145. https://doi.org/10.1111/ajps.12089

Messing, S., & Westwood, S. J. (2014). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, *41*(8), 1042–1063. https://doi.org/10.1177/0093650212466406

Mialon, H. M., & Rubin, P. H. (2008). *Economics, Law and Individual Rights*. Routledge.

Mutz, D. C. (2002). Cross-cutting Social Networks: Testing Democratic Theory in Practice. *American Political Science Review*, *96*(1), 111–126. https://doi.org/10.1017/S0003055402004264

Pariser, E. (n.d.). *Beware online “filter bubbles.”* Retrieved from https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles

Phillips, J. (2018). Risk in a digital age: Understanding risk in virtual networks through digital response networks (DRNs). *International Development Planning Review; Liverpool*, *40*(3), 239–272. http://dx.doi.org/10.3828/idpr.2018.18

Rowland, F. (2011). The Filter Bubble: What the Internet is Hiding from You (review). *Portal: Libraries and the Academy*, *11*(4), 1009–1011. https://doi.org/10.1353/pla.2011.0036

Sears, D. O., & Freedman, J. L. (1967). SELECTIVE EXPOSURE TO INFORMATION: A CRITICAL REVIEW. *Public Opinion Quarterly*, *31*(2), 194–213. https://doi.org/10.1086/267513

Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review*, *34*(3), 150–160. https://doi.org/10.1177/0266382117722446

Stroud, N. J. (2008). Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure. *Political Behavior*, *30*(3), 341–366. https://doi.org/10.1007/s11109-007-9050-9

Sun, T., Zhang, M., & Mei, Q. (2013). Unexpected Relevance: An Empirical Study of Serendipity in Retweets. *Seventh International AAAI Conference on Weblogs and Social Media*. Presented at the Seventh International AAAI Conference on Weblogs and Social Media. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6129

Technology | Academics | Policy—Polarization in 2016. (n.d.). Retrieved November 4, 2019, from http://www.techpolicy.com/Blog/March-2016/Polarization-in-2016.aspx

The Journal of Politics on JSTOR. (n.d.). Retrieved November 4, 2019, from https://www-jstor-org.mutex.gmu.edu/journal/jpolitics?refreqid%3Dexcelsior%253Aff899084a3c11701b4580e81081566ef

Wong, J. C., Levin, S., & Solon, O. (2016, November 16). Bursting the Facebook bubble: We asked voters on the left and right to swap feeds. *The Guardian*. Retrieved from https://www.theguardian.com/us-news/2016/nov/16/facebook-bias-bubble-us-election-conservative-liberal-news-feed