News-Related Social Media Use,
Political Knowledge, and Participation in the 2016 Election

By
John Roberson

Senior Honors Thesis
Department of Public Policy
University of North Carolina at Chapel Hill

April 26, 2019

Approved:

Douglas Lee Lauen, Thesis Advisor

Rebecca Kreitzer, Second Reader

Abstract

Social media's role in the 2016 election was one of many aspects of that election which make it unique. Record numbers of Americans were active social media users in 2015-2016, and many used the platforms not only for entertainment, but to learn about the news. Social media were also used by candidates in novel ways and to unprecedented extents, and the quality of information distributed through social media came into question. Previous studies have shown that news-related social media use leads to increases in political knowledge and likelihood of voting, but the unique circumstances surrounding the 2016 election make generalization of those findings to that election perilous. Therefore, this thesis used national survey data from the 2016 Cooperative Congressional Election Survey to investigate the relationships between news-related social media use, political knowledge, and participation specifically in the 2016 election. Exploratory factor analyses, logit regressions, and marginal effects were used. Analysis found that consuming social media news content did predict higher likelihood of voting in 2016, and it found that that effect was partially mediated by an increase in political knowledge. Additionally, evidence suggested that whether someone shared, commented on, or forwarded news-related content – in addition to consuming it – was found to moderate each of the aforementioned relationships, though that finding was not statistically significant. The relationships between social media use, knowledge, and voting also varied by state and by party. Further research should consider personality in analysis and use experimental, rather than observational, methods to the extent possible.

TABLE OF CONTENTS

INTRODUCTION **1**
1. Web 2.0, Social Media, and a Shift in the Flow of Information 2
2. Policy Relevance 3
3. The 2016 Election 6
4. Key Research Questions 8
5. Thesis Roadmap 9

BACKGROUND **10**
1. Social Media Use 10
2. Political Knowledge 13
3. Political Participation 19
4. Putting It All Together 20

METHODS **27**
1. Causal Diagram 27
2. The 2016 Cooperative Congressional Election Survey (CCES) 29
   2 A. Vote Validation in the CCES 31
3. Modeling Social Media Behavior 37
4. Modeling Political Knowledge 44
5. Other Selected Variables for Analysis 48
6. Regression Models 50

RESULTS **53**
1. Descriptive Analysis 53
2. Regression of Voting on Social Media News Use 59
3. Regressions of Political Knowledge on Social Media News Use 69

CONCLUSIONS **72**
1. Evaluation of Hypotheses 72
2. Limitations of Methods 74
3. Implications of Findings 75
4. Suggestions for Future Research 77

REFERENCES **80**

APPENDIX A: FACTOR ANALYSIS OUTPUT **92**

APPENDIX B: LOGIT REGRESSION TABLES **96**

TABLE OF TABLES

Table 1: Determinants of Political Knowledge 16
Table 2: Summary of Potential Confounding Variables 26
Table 3: Defining Three Validated Voter-Turnout Measures 34
Table 4: Selecting a Validated Voter-Turnout Measure 36
Table 5: Social Media Behavior EFA Results 42
Table 6: Political Knowledge EFA Results 47
Table 7: Descriptions of Voting Regression Models 51
Table 8: Weighted Descriptive Statistics of Key Variables 53
Table 9: Exponentiated Coefficients on Treatment and Mediating Variables, Logit Regressions of Voting on Social Media Use 60
Table 10: Frequency Distribution of Social Media Use Response Patterns 92
Table 11: Social Media Behavior Preliminary EFA Results A 93
Table 12: Social Media Behavior Preliminary EFA Results B 93
Table 13: Social Media Behavior EFA Results - No Omitted Loadings 93
Table 14: Frequency Distribution of Political Knowledge Response Patterns 94
Table 15: Political Knowledge Preliminary EFA Results A 95
Table 16: Political Knowledge Preliminary EFA Results B 95
Table 17: Political Knowledge EFA Results – No Omitted Loadings 95
Table 18: Full Logit Regressions of Voting on Social Media, Mediated and Unmediated, with and without State Fixed Effects 96
Table 19: Fractional Logit Regressions of Political Knowledge on Social Media News Use..... 99

TABLE OF FIGURES

Figure 1: The Mediated Relationship Between Social Media News Consumption and Voting.. 28
Figure 2: Social Media EFA Loadings, Before and After Rotation 41
Figure 3: Kernel Densities of SM Factor Scores 43
Figure 4: Age Distributions by Model Sample 55
Figure 5: Distribution of Social Media Consumption by Model Sample 56
Figure 6: Distribution of Social Media Engagement by Model Sample 56
Figure 7: Distribution of Personal Knowledge Scores by Model Sample 56
Figure 8: Distribution of General Knowledge Scores by Model Sample 56
Figure 9: Distribution of Educational Attainment by Model Sample 57
Figure 10: Income Distribution in Analytical Sample 57
Figure 11: Ideological Distribution by Model Sample 58
Figure 12: Distribution of Party Identifications by Model Sample 58
Figure 13: Party IDs of Self-Reported Ideological Moderates 59
Figure 14: Odds Ratios Across Mediated Models of Voting (Comparing Methods for Calculating Political Knowledge) 61
Figure 15: Predictive Margins of Consumption on Voting, by Level of Active Engagement 62
Figure 16: Contrasts of Predictive Margins of PersonalKnowledge on Voting 64
Figure 17: Heatmaps of Statewide Turnout and State-Level Contrasts in Predicted Turnout by Level of Social Media News Consumption 65
Figure 18: Scatterplot of Statewide Turnout against State-Level Contrasts in Predicted Turnout by Level of Social Media News Consumption 66
Figure 19: Odds Ratios Across Split Models of Voting by Party ID (Comparing Models Limited to only Democrats and only Republicans) 67
Figure 20: Predictive Margins of Social Media News Consumption on Voting by Level of Active Engagement, across Separate Models for Democrats and Republicans........ 68
Figure 21: Predictive Margins of Social Media News Consumption on Personal Political Knowledge, by Level of Active Engagement 69
Figure 22: Predictive Margins of Age on PersonalKnowledge Score 70
Figure 23: Matrix of Predictive Margins of Personal Political Knowledge by Party ID and Ideology 71

Introduction

Social media is pervasive in many Americans' daily lives in 2019. To some, it is a way to connect with old friends; to others, a way to catalog their daily experiences; and to many others, something different. One marked change in social media over the last decade is the development of social networking sites as platforms for the dissemination of news. Traditional news outlets have inserted themselves into the social media sphere and are one source of the news information that is spread across social networking sites, but such sites are also part of Web 2.0 at their core – and as such, any site member is empowered to be a source of information as well, with varying degrees of influence based on the size of their networks. People and media companies share all kinds of news over social media, but whenever "news" is referred to in the rest of this thesis, it should be assumed that the term refers to political news specifically.

This thesis looks at the role(s) that social media news played in the 2016 elections. That election cycle saw a rising trend in general popularity of social media, novel use of social media by the Republican nominee, interference by foreign interests through social media, social media-based propaganda loops, and other trends that continue to be identified and investigated. People both consumed and engaged with news content over social media, and it can be expected that what happened there may have impacted their behavior on Election Day. In the remainder of this chapter, I quickly review the development of Web 2.0 and social media, the democratization of information, and potential implications for society. At the end of the chapter, I introduce the key research questions that guide my analysis and I provide a roadmap for the rest of this thesis.

1. Web 2.0, Social Media, and a Shift in the Flow of Information

Until the 1980s, information mainly was spread one of a few ways: through mass media, through education, or by word of mouth. In that traditional media context, word-of-mouth information spread was limited, and information flowed in society unidirectionally, flowing from a few mass media producers to mass consumers (Mandiberg 2012, 1). Starting in the 1980s, technologies like photocopiers, home video cameras, and home computing began allowing one person's information to reach more people without going through mass media (1). Soon after that came a boom in home computing, which brought mass access to the internet and dramatically increased people's connectivity with the outside world. Still, that connectivity was limited.

It can be said that the internet so far has developed through two eras (and is arguably about to enter a third). A 2012 blog post from a web design studio provides an oversimplified, but useful way to distinguish the eras: in computer science language, Web 1.0 was the "readable" era of the internet, Web 2.0 is the "writable" phase, and, while not really realized yet, Web 3.0 will be the "executable" phase of the internet (WittyCookie 2012). In more detail, "Web 1.0," was mostly a one-way street. Sites made information available to users, but users had little to no opportunity to return the favor. In Web 2.0, the internet is a two-way street: sites provide users with information, users make use of it and provide information back to the sites, sites update their services based on the received data, and the cycle continues (O'Reilly 2007). Importantly, this process introduced the possibility of not only making one user's data available to the website, but to other users of the site as well – the birth of social media (Mandiberg 2012, 2).

Now, in contrast to the rigidly-unidirectional information flow of the pre-1980s, the pattern in modern and social media environments is often turned on its head to be bottom-up, as well as laterally-oriented (Buckingham 2015; Montgomery & Gottlieb-Robles 2006). David

Buckingham, a giant of turn-of-the-century media literacy scholarship, describes this very clearly, writing about the general consensus that

> the interactive, participatory possibilities of digital media are believed to transcend the limitations of hierarchical, top-down 'mass' media and hence to undermine what are seen as the authoritarian 'knowledge politics' of traditional pedagogy. The potential they offer for learners to become creators of knowledge – rather than merely 'consumers' has been seen by some as little short of revolutionary. (Buckingham 2015, 9)

Jay Rosen (2007) describes a similar phenomenon, writing a fictional open letter to media producers in the voice of "the people formerly known as the audience." In that letter, he tells mass media "You don't own the eyeballs. You don't own the press, which is now divided into pro and amateur zones. You don't control production on the new platform, which isn't one-way. There's a new balance of power between you and us" (15).

2. Policy Relevance

"Welcome to the second decade of the 21st century," Ernest Morrell writes in 2012, "where information has been globalized, digitized, and sped up to move at the speed of thought" (300). The Pew Research Center estimates that nine in ten Americans use the internet in 2018, up from only half of the population in 2000 (Pew 2018a). There are certainly exceptions, groups that are less connected than the average (primarily older, rural Americans), but even those disparities are rapidly disappearing (Pew 2018a).

Americans are using the internet to get on social media. In 2008, roughly two in ten Americans used social media; by 2010, that had doubled to just shy of four in ten (Hitlin 2018; Moore & McElroy 2011, 267). By, 2016 the figure rose to almost seven in ten (Hitlin 2018).

Certain social media platforms are more popular than others. By far the most popular is Facebook: 68% of Americans report using that platform, compared to 35% using Instagram, 29% using Pinterest, 27% using Snapchat, 25% using LinkedIn, and 24% using Twitter (Smith & Anderson 2018, 2). Facebook has been the most popular site in every Pew survey since the Center began asking about respondents' use of specific social media sites in 2012, always being used by roughly three times as many people as the next-most popular site (2).

These people not only "use" the sites; most use them daily. 74% of Facebook users report using the site at least once a day, and two thirds of those report using it more than once a day. That amounts to over a third of Americans using Facebook "several times a day," a figure that did not change significantly from 2016 to 2018 (4-5). The shares of Snapchat, Instagram, and Twitter users who use those sites daily are smaller than for Facebook, but still significant: roughly 60% for Snapchat and Instagram and 46% for Twitter (4).

Additionally, Americans do not silo themselves onto only one of these platforms. Just shy of three-quarters of Americans use more than one platform, and the median American uses three (5). Lastly, although more people would find it "not hard to give up" social media than would find it "hard" (59% versus 40%), the share of users who would find it hard has grown over recent surveys, from 28% in January 2014 (6).

All of this social media use is not purely for entertainment. Before the 2008 presidential election, political organizing and news-sharing over the most popular social media site, Facebook, was relatively scarce. Even with 500 million users worldwide in 2008, only four years removed from its launch, Facebook was still seen as little more than a social networking site used to share personal information and life updates. Woolley, Limperos, and Oliver (2010) argue that it was the 2008 election that saw social media transition from being "virtually unknown in the realm of politics to a budding form of political communication" (632). Contemporary traditional media outlets and scholars alike were struck by the role of social media (namely, Facebook) in the election; Woolley, Limperos, and Oliver (2010) highlight one CNN headline that questioned "Will the 2008 presidential election be won on Facebook?" (632).

Since 2008, receiving news content via social media has been on the rise. A 2017 Pew Research Center study found that two-thirds of Americans get at least some of their news on social media, with one in five Americans doing so "often" (Shearer & Gottfried 2017, 2). Much of this growth was driven by an increase in older, less-educated, and non-white Americans getting news from social media (2). For example, over half of Americans aged 50+ reported getting at least some of their news from social media sites for the first time since the question first asked in Pew surveys in 2013 (2, 14-15). For Americans under the age of 50, the share of people who got news via social media was over 75% (2).

Any government has an intrinsic interest in knowing how information spreads though its society, and that interest is even stronger when the outcome in question (civic participation) is the foundation of the system of government. With over 90% of Americans online and consuming massive amounts of information each year, it is important to understand how that information might affect their decision to vote.

Each successive development in media and information technology in the history of the United States has brought concerns about its impact on American civil society and civic life (Allcott & Gentzkow 2017). The era of "yellow journalism" is well documented as a period in which media sensationalism led to distinctly undesirable outcomes (Kaplan 2002). Radio and television led to their own unique problems, with critics (correctly) worrying that the platforms would cause a problematic distillation of information into "sound bites," concentration of power into the hands of large media companies, and increased national preference for more attractive, gregarious candidates (Lang & Lang 2002, Bagdikian 1983). The introduction of internet media brought the novel concern of echo chambers; when people are given a wide set of possible news sources to choose from, scholars were concerned that people would self-select into clusters of people already more like themselves (Sunstein 2007). Now, in the social media age, a new concern has arisen, in addition to the existing ones, which deals with the extent of the average person's influence on others: most people can make and share content in the blink of eye, some with a similar reach to many traditional media sources just half a century ago (Morrell 2012; Gil de Zúñiga et al. 2014; Allcott & Gentzkow 2017).

3. The 2016 Election

The 2016 election was exceptional in many regards, and since November 2016, academics have worked to understand that election and the forces that shaped it. A plethora of perspectives on that election have been put forward, and the field of thought only continues to expand. Daniel Kreiss (of UNC's Media and Journalism School) reviewed three books that lay out three of the leading perspectives on the 2016 election: that hackers (foreign and domestic) had great potential to set the national press agenda (Jamieson 2018); that changing demographics, a fractured Republican Party, and "explicit racial appeals" made by candidate Trump combined to radicalize the election and activate key voting blocs (Sides, Tesler, & Vavreck 2018); and that "Fake news entrepreneurs, Russians, the Facebook algorithm, and online echo chambers" led to a "crisis of disinformation and misinformation" primarily in the right media ecosystem (Benkler, Faris, & Roberts 2018, 11, 15). Benkler, Faris, and Roberts (2018) provide evidence of meaningful differences between left- and right-wing media ecosystems, finding that while both sides see "supply and demand of false partisan narratives," the left includes not only left-slanted media but centric media and slanted media that nonetheless adhere to journalistic norms, while the right-slanted "ecosystem is both more isolated from the center and there are few journalistic breaks on falsity" (Kreiss 2019, 3).

One common thread through each of these perspectives is social media. Hackers, "fake news entrepreneurs," foreign interests, and simple partisans populated echo chambers, which directly influenced social media users and also often served as source material for (mostly right-wing) which picked up the content in a "propaganda feedback loop." Social media also did not only benefit the non-traditional, populist candidate who ultimately won the US presidency, it also played a significant role in the achievements of an unsuccessful populist candidate (Bernie Sanders) in the Democratic primary (Groshek & Koc-Michalska 2017). The full extent of social media's impact in 2016 continues and will continue to be hashed out, but it appears that there is consensus that it did have an important effect.

Importantly, this thesis does not distinguish between "real" and "fake" news, one of the social media phenomena most-frequently discussed in relation to the 2016 election. The reason is simple – the data used in analysis provided no means to evaluate the veracity of information obtained through social media. Nonetheless, the "fake news" phenomenon – whether foreign governments, political partisans, or just internet "trolls" manipulated and polluted news content on social media – merits extended discussion in other work. The full extent to which this phenomenon influenced the 2016 election may never be known. One study by BuzzFeed news found that "the top-performing fake" news stories on Facebook in the four months preceding the election received more shares, reactions, and comments than the top-performing stories from mainstream news outlets (Silverman 2016). However, further analysis has called those findings into question (Rinehart 2017), illuminating how difficult it is to consider "fake news" with robust methodology. As already mentioned, making the distinction between "real" and "fake" news is beyond the scope of this thesis. Nonetheless, knowing the extent to which people's voting is affected by *any* news they consume over social media – valuable for policymakers and social media providers in its own right – can also give insight to the possible extent of any effect of "fake" news that might have been consumed as part of that.

4. Key Research Questions

In a media environment where Americans can get news not only from television, the newspaper, or talk radio, but also across social media, where it may be more or less filtered, personalized, biased, or detailed, what effect does that news have on Americans' likelihood to participate in the political process? And importantly, does getting news information from social media make people more knowledgeable about politics, which one would hope that people participating in the political process would be?

Based on previous work in this field, the following hypotheses were formulated:

H1: Consuming news through social media was positively related to voting in 2016.

H2: The effect of social media news use on voting in 2016 was partially mediated by an increase in political knowledge.

H3: News-related social media use will have affected Democrats and Republicans differently in 2016.

Additionally, I formulated two further hypotheses dealing with questions of measurement – the reason for this consideration is explained in Chapter Two:

H4: Measures of political knowledge which are limited only to actual attempted answers will be subject to less bias on the basis of sex.

H5: Measures of political knowledge which ask respondents about their own political representation, rather than national political contexts, will be subject to less bias on the basis of sex.

5. Thesis Roadmap

This thesis is organized into four remaining chapters. In the next chapter, I examine past research, with a primary focus on uncovering potential confounding factors that affect two or more of: consumption of social media news content, political knowledge level, and likelihood to vote. I also review a few papers that have undertaken research questions similar to those posed by this thesis. In Chapter Three, I construct a causal diagram to explain the relationships between social media use, political knowledge, likelihood to vote, and the confounds uncovered in Chapter Two. Then, I introduce the data used, the 2016 Cooperative Congressional Election Survey (CCES) Common Content. I review preprocessing steps taken and explain two exploratory factor analyses that were performed to obtain scores for social media use and political knowledge. Chapter Three concludes with the introduction of regression models, which apply the operationalized variables from the previous sections to the causal diagram from the beginning of the chapter. In Chapter Four, I present descriptive analysis of the variables used, followed by the results of the regression analyses. Finally, in Chapter 5, I summarize the findings of analysis and conclude by drawing wider conclusions.

Background

In order to isolate the relationship between social media news use and voting, all plausible factors linking the two must be uncovered. After adding in political knowledge, all plausible factors linking it to either or both previous two must be explored as well. In this chapter, I undertake that exploration in three parts, documenting the predictors of each variable independently. At the end of the chapter, I review the literature that has previously investigated the places where those predictors overlap.

1. Social Media Use

Previous scholarly work that attempts to predict social media use falls into three main categories: psychological, network-based, and demographic. Because the CCES does not contain any psychological or social-network questions, the literature in those two categories is of limited assistance to this thesis. Nonetheless, I briefly review the highest points of the psychological papers below and extract a few useful notes. Then, I review in greater depth the papers that examine social media use from a demographic perspective.

Psychologists generally agree that personality predicts a considerable amount of social media behavior (Moore & McElroy 2011; Correa, Hinsley, & Gil de Zúñiga 2010; Ross et al. 2009). When discussing "personality," contemporary psychologists frequently use the Five Factor Model to provide structure to their work, according to Moore and McElroy (2011, 268). The Five Factor Model delineates personality along five axes: extraversion, agreeableness, conscientiousness, emotional stability, and openness to new experiences.

However, other scholars, such as Khang, et al. (2014), explain social media use purely as a function of habit, rather than any other psychological motivator. Notably, Khang et al. (2014)

employ exploratory factor analysis to create social media scales, which supports the methods of this thesis. Nonetheless, habits must be formed, and Khang, et al. (2014) offer little to explain who develops social media use habits. The Five Factor personality model might explain how people enter into social media habits, but demographic factors may explain part of it as well. Scott et al. (2017), which is explored below as part of the category of papers using demographic predictors of social media, focuses in particular on the characteristics of those social media users who "habitually engage" (312).

Moore & McElroy (2011) of the psychological category present one finding that is useful in establishing explanatory variables that will be included in this thesis: they find evidence that gender is a strong predictor of Facebook usage and content, and they suggest that it should be controlled for in future work (272). Demographically-focused papers that are cited below extend this effect from Facebook to all social media platforms.

However, as stated before, the majority of the psychological school's findings do not factor into this thesis. This is ultimately a constraint imposed by the observational design of this thesis; the CCES simply does not ask any questions that could be used to evaluate respondents' personalities in any robust way. It does ask personal opinion questions, e.g. what the respondent thinks government priorities should be, but none would translate cleanly into psychological models. Additionally, demographic controls will not control for personality; the most predictive personality traits, primarily extraversion, have been found to be relatively randomly distributed among demographic groups (Goldberg, et al. 1998, 399). Ultimately, differences in personality must be left in residuals in this thesis' analysis.

The more-relevant class of papers that attempt to explain patterns in social media use are those that do so using demographic factors like age, education, gender, race, and so on (Scott et al. 2017, Duggan & Brenner 2013, Lenhart, et al. 2010). This thesis relies mostly on the work of Scott et al. (2017), which attempted to create profiles of social media users among "emerging adults," people aged 18-26. Their work contributes several explanatory variables for social media use, as well as introduces a framework for considering "active" social media use, a distinction similar to the consumption versus active engagement dynamic in question in this thesis.

In their methods, Scott et al. (2017) distinguish between two dimensions of social media use: frequency and engagement (312). "Frequency," which according to Scott et al. (2017) is the focus of "the vast majority" of previous studies in the field, entails minutes of use, frequency counts of checking feeds, and so on. "Engagement," on the other hand, entails responding to other users, (re)posting content, and filtering content by blocking, unfriending, or unfollowing. This "participative," habitual engagement through social media is the active social media behavior that this thesis uses to distinguishes from basic social media news consumption.

In their results, Scott et al. (2017) find that gender, race, and education are all statistically significant predictors of social media use, though their effects are split between frequency and engagement (316, Tables 5-6). Scott et al. (2017) find that women make up a statistically-significantly greater portion of high-frequency users than low-frequency users, but find that – among people who log on at roughly equal frequencies – gender did not affect the extent to which they engaged with content by interacting with other users, posting, and so on. However, Scott et al. (2017) did find that engagement was associated with both race and education, with highly-engaged users tending to be white and more highly-educated.

Scott et al. (2017) found income and age to be statistically insignificant (316). However, Scott et al.'s (2017) findings regarding income and age should be taken with a grain of salt, as their sample consists of people aged only 18-26. Both age and income will vary far more in an age-unlimited sample than in a sample limited to people in early young adulthood. Therefore, both age and income could very possibly be significant predictors of social media use in the wider American public even though Scott et al.'s (2017) findings suggest otherwise.

Evidence to that fact is provided by multiple surveys conducted by the Pew Research Center. Smith and Anderson (2018) find "substantial differences in social media use by age" (3). Specifically, they find that social media use and age are negatively related: the older someone is, the less likely they are to use social media. 88% of those aged 18-29 use some form of social media, compared to 78% among 30 to 49 year-olds, 64% among 50 to 64 year-olds, and 37% among those aged 65 and over. Shearer and Gottfried (2018) find that 45% of American adults over the age of 50 got news from social media sites in a survey conducted in 2016, while 78% of Americans ages 18-49 got news from social media. Shearer and Gottfried's (2018) findings also support education as a predictor of social media news use: 60% of Americans with some college experience or less education reported getting news from social media sites in that same survey, compared to 68% of college graduates and higher who reported doing so.

2. Political Knowledge

The next set of predictors that must be identified are those of political knowledge levels. First, however, a distinction must be made about to what the term "political knowledge" refers. Scholars have employed myriad measures for political knowledge. Most of them are based on surveys, but there is significant variation in the types of questions asked (Pew 2018b, Freiling 2017, APPC 2014). Political knowledge survey questions can be grouped into three main families: identification questions (Pew 2018b, Freiling 2017, APPC 2014), civic questions (Pew 2018b, APPC 2014), and factual questions (Pew 2007). The most common are identification and civic questions; those are explored in greater detail below. Factual questions are less common, perhaps due to their limited "staying" power. For example, the 2007 Pew survey asked respondents to estimate the number of U.S. troops who had been killed in Iraq at the time[1] and to identify the subject country of "many recent news stories about tainted food and dangerous toys[2]" (6-8). Such questions are useful for evaluating knowledge of "hot topics," but because hot topics by definition come and go, using them to evaluate trends is, relatively, more difficult than identification and civic knowledge, whose questions remain essentially constant.

Civic questions ask respondents about structures, rules, and history of American government. For example, a 2018 Pew survey categorized respondents as having either "High," "Medium," or "Low" civic knowledge based on their answers to four questions: who breaks tied votes in the U.S. Senate, how many votes are needed to end a filibuster in the Senate, which amendment to the Constitution sets presidential term limits, and what the Electoral College is (Pew 2018b, 117). A 2014 survey by the Annenberg Public Policy Center (APPC) asked similar questions, such as the vote threshold for presidential veto overrides, the names of the three branches of federal government, and the degree of federal and legislative oversight over the Supreme Court (APPC 2014, 1). Performance on these questions varies. Around a third of Americans could name all three branches of government in the 2014 APPC survey, but also around a third could not name even one (1). Only a quarter of Americans knew that two-thirds of both the House and Senate must vote to override a presidential veto (1). On the Pew survey, only 23% of Americans answered all four questions correctly; 44% answered two to three questions correctly, and 32% were correct on one or none (Pew 2018b, 117).

[1] Around 3,500
[2] China

Identification questions ask more contemporary questions about the state of politics: for example, can the respondent correctly identify which party is in control of a given legislative body? Brennan cites such a statistic in his book, claiming that "citizens generally don't know which party controls Congress" (Chapter 2, 2). Numerous surveys put this claim to the test, and the results are mixed. The 2014 APPC survey supported Brennan's claim, finding that 38% of Americans correctly identified the party in control of the House[3] at the time; the same percent did so for the Senate[4,5] (APPC 2014, 2). A 2018 Pew Research Center study contradicted Brennan's claim, however: 82% of its survey respondents correctly identified the Republican majority in the House and 83% identified the Republican majority in the Senate. 75% correctly identified both (Pew 2018b, 108). There are many potential explanations for this difference, including difference in time frame (2014 vs. 2018) or sampling bias. It is unlikely for such a large difference to be the result of change over time, but the changed nature of politics after the 2016 election may contribute to the difference.

When asked to identify the party of their personal Representative in Congress, Americans typically answer correctly about half the time. A 2017 study found that 56% of respondents claimed to know the party of their district's current member of Congress (Freiling 2017). It is worth noting that their claim was not validated; the 56% figure represents *claims* of knowledge, not verifiable knowledge. Therefore, those figures should be taken with a grain of salt.

When it comes to predictors of political knowledge, Brennan uses identification-type questions (as opposed to civic or factual questions). He identifies eight specific axes along which

---

[3] The Republicans
[4] The Democrats
[5] For both chambers, over 40% of respondents did not even guess, saying they did not know. That figure was up from just over 25% in 2011 (APPC 2014, 2).

such political knowledge varies, which are summarized in **Table 1** below. According to his findings, political knowledge is significantly associated with many fundamental demographic factors, ranging from the relatively more-malleable, e.g. education and geographic region, to the immutable, e.g. age and race.

Table 1: *Determinants of Political Knowledge*

Characteristic	***Knowledge Positively Correlated With…***	***Knowledge Negatively Correlated With…***
Education	Having a college degree	Having a H.S. diploma or less
Income	Earning above-average income; also being in top 25% of income earners	Earning below-average income; also being in the bottom 25% of income earners
Region of the US	Living in the West	Living in the South
Partisanship	Being or leaning Republican	Being a Democrat or leaning Independent
Age	Being middle-aged (35-54)	Being any other age
Race	-	Being black
Gender	-	Being female

SOURCE: Brennan "Against Democracy" 2016, Chapter 2, Page 7

A 2018 Pew Research Center survey supports several of Brennan's claims (Pew 2018b, 110). It found a positive, linear relationship between educational attainment and "civic knowledge," suggesting that someone with a postgraduate degree is almost four times more likely to have "high" civic knowledge than someone with a high school diploma or less. It also found that age was associated with civic knowledge, though it observed a positive, linear relationship in contrast to the parametric relationship posited by Brennan (108).

That Pew survey also makes several findings that are in total contrast to Brennan's findings. For example, the Pew survey found "no major differences" in civic knowledge by party identification (110). Almost exactly the same percent of Democrats were found to have high civic knowledge as Republicans; the same was true for having medium civic knowledge, as well as low (109). Instead, it found that the respondent's ideology was a significant predictor. People who identified farther from the center within their party were more likely to have high civic knowledge (e.g., conservative Republicans were more knowledgeable than moderate or liberal Republicans) (110).

An important note must be made regarding common measures for political knowledge. First, in his book, Brennan raises concerns about the fundamental accuracy of approximations of people's knowledge levels – not only self-reported ones like in Freiling (2017), but also calculated ones like those used in Pew (2018b) and this thesis. Naturally, people may simply be wrong about knowing something, as is the concern in Freiling (2017). More critically, on calculated political knowledge measures, such as whether someone correctly identifies the party in control of a legislative chamber, some error can be introduced due to guessing; there is some non-zero, relatively significant change that someone simply guessed the correct answer (Brennan 2016, chapter 2, page 3). This is of concern for this thesis, as the political knowledge measure that was used in analysis is a calculated political awareness measure.

The second note is an extension of Brennan's concern about people's ability to guess on political knowledge questions: multiple scholars have found a distinct gendered difference in likelihood of such guessing on political knowledge questions (Delli Carpini & Keeter 1996, Verba, Burns, & Schlozman 1997, Lizotte & Sidman 2009, Kenski 2006, Miller 2018). Men have been found to be more likely to guess on political knowledge questions, while women are more likely not to attempt the question if they don't know the answer, instead responding "Don't Know" or "Not Sure" – if given the option. Lizotte and Sidman (2009) found that controlling for men's higher likelihood to guess reduces the knowledge gender gap by around 36%. That effect has been found to persist even after controlling for factors such as age, race, education, income, political engagement, and employment (Garand, Cuynan, & Fournet 2004).

One explanation for this persistent effect has been the nature of questions asked. Many authors have found that broad questions about "national-level politics and the rules of the game" favor men (Miller 2018, 177). In a 2011 paper, Kathleen Dolan finds that asking gender-relevant questions about politics, e.g. the percent of women in Congress, leads the long-reported gender differences in political knowledge to essentially "disappear." Other topics on which women tend to perform equally with or better than men include identifying candidate positions on abortion and naming the head of the local school board (Delli Carpini & Keeter 1996, Shaker 2012).

Miller (2018) references literature providing two theoretical explanations for women's lower performance on national, institutional questions but higher performance on gender-relevant or local questions. First, Dolan (2011) found that people are more likely to pay attention to politicians who are descriptively more like them – and because there are far fewer women in national public office than men, women may pay less attention to national politicians, because they relate to them less. Second, Stolle and Gidengil (2010) found that women benefit more frequently from government services and are employed in government more frequently than men, yet questions gauging knowledge of government services are almost never included in traditional political knowledge questionnaires.

Unfortunately, there are no gender-relevant questions of the type described by Dolan (2011) in the CCES. However, one subset of political knowledge questions in the CCES may be less gender-biased than the other, bigger-picture questions: the CCES includes four questions that ask respondents to identify their own personal representatives in government, rather than identify large, national balances of power. Miller (2018) offers some theoretical foundation to expect that these questions may be less gender-biased. Her research finds that there is no statistical difference between the rate at which men and women correctly identify the majority party in the respondent's home state (Table 3) – a more local identification, like the four personal representation identification questions. This topic is further explored in Chapters Three and Four when this thesis' specific knowledge measures are created and explored in descriptive analysis.

Fortin-Rittberger (2016) offers an additional step that can be taken to account for some of the gender bias in survey-response based political knowledge scores; her findings suggest that calculating scores as a proportion of correct responses out of only valid responses removes a nontrivial amount of the gender difference in political knowledge (Table 2). Considering respondents' accuracy only if they affirmatively answer removes a substantial amount of gender's effect on knowledge, by accounting for the gendered difference in likelihood of guessing versus abstaining from answering (393). Such a metric will be used in this thesis' analysis. Nonetheless, gender will still be controlled for, because of its effects on social media and on voting, as well as because traditional-type political knowledge questions must be used.

3. Political Participation

The final set of predictors to identify are those of the outcome: voting. This section relies mostly upon the work of the MIT Election Data and Science Lab, which used data from the United States Election Project and the Census Bureau to calculate nationwide 2016 voter turnout by demographic groups. Their analysis begins by touching on over-reporting bias in survey-based voter turnout rates, which is addressed in this thesis as well, in Chapter Three.

Then, the MIT lab provides subpopulation means of voter turnout for education, income, age, marital status, sex, and race/ethnicity. Nationwide, the MIT lab estimates that 77% of people with more than a high school degree voted in 2016, while only 44% of people with a high school degree or less did. Another wide gap is found when stratifying by income: people in families with income of more than $50,000 were 69% likely to vote, while people in families with income of less than $50,000 were 50% likely to vote. Age also was a significant predictor of voter turnout; those aged 60+ voted at a rate of 72%, while those aged 31 to 60 voted 62% of the time and those aged 18 to 30 voted only 44% of the time. Married people voted 18 percentage points more frequently than unmarried people (69% versus 51%) and women voted five percentage points more frequently than men (63% versus 58%). Finally, white people were the most likely racial/ethnic category to vote, at 65%, followed by black people at 60%. Hispanic, Asian-American, and "all other" people voted at roughly 46%. All these turnout rates also varied by state. I speak in more detail about state-by-state turnout in Section 2.A below, but speaking generally, state turnout rates in 2016 ranged from 42% to 74%, with a mean of 61% and a standard deviation of 6%.

4. Putting It All Together

The three previous sections have reviewed literature that describes the individual predictors of each of the three larger concepts at hand in this thesis: social media (news) use, political knowledge, and political participation. This section reviews the existing literature that has connected two or more of those. Much of that work comes, in one way or another, from Homero Gil de Zúñiga of the University of Vienna. After reviewing that literature, the chapter concludes with a summary table of potentially confounding variables in analysis.

Going back to the mid-2000s, not too long after the introduction of Web 2.0, two papers looked at the effect of informational internet use on political knowledge and participation. Both papers found that engaging with political content online impacted both knowledge and participation. Kenski and Stroud (2006) found that internet access and exposure to online political content were statistically significant, though substantively limited, predictors of political knowledge and participation. In addition, they found that sex, age, race, education, income, and partisanship – both direction and strength – were predictors of *both* outcomes (knowledge and participation) (Table 3).

Shah, Cho, Eveland, and Kwak (2005) looked at information *seeking,* which is slightly different from Kenski and Stroud's exposure. Additionally, while Kenski and Stroud (2006) control for gender, race, education, and income, Shah et al. (2005) do not use any demographic controls in their models. Shah, et al. developed three types of structural equation models, one cross-sectional, one with fixed effects, and one auto-regressive. Because this thesis does not deal with panel data, the cross-sectional model is most comparable and will be the most discussed, though the patterns in Shah et al.'s findings do not differ greatly among the three models (551).

In their model, Shah et al. measure three types of information seeking: searching online, reading the newspaper, and watching TV. They also qualify two types of political expression: "interpersonal political discussion," i.e. talking about politics with friends, family, neighbors, coworkers, etc., and "interactive civic messaging," e.g. emailing a politician or writing a letter to the editor (541). Finally, they measure civic participation as self-reported frequency of volunteering, working on community projects, attending community meetings, or doing otherwise charitable work (540).

In the cross-sectional model, online information seeking was found to strongly, positively predict both types of political expression (545-546). In turn, both types of political expression were found to positively predict civic participation. However, they also note that when all paths are held open, i.e. for all three information seeking types (online, newspaper, TV), only reading the newspaper is found to have a direct effect on civic participation (546). Nonetheless, the indirect effect of online information seeking on civic participation through the two types of political expression remains.

The model structure and findings of Shah et al.'s paper are both very informative for this thesis. Nevertheless, the connections are not perfectly lined up. Given the time-frame of the paper's analysis, the 2000 election, the types of online activity measured here are starkly different from the online social media news use at hand in this thesis. In 2000, if people wanted information about politics, they truly had to *seek* it out, which is why Shah et al. term it "information-seeking" in their paper. That type of behavior is also fairly similar to reading the newspaper, which requires a conscious choice to obtain the information and then be able to read it. The same could be said of watching TV news, though the comparison is less strong.

The social media news use that is being studied in this thesis requires not nearly as much "seeking-out" of the information. People may encounter news items on their social media feeds because they follow an outlet's page, because a personal connection shared it, or even because it was advertised. The amount of effort required to encounter news on social media in 2016 was very different from the threshold set in Shah et al.'s analysis.

Additionally, the civic participation outcomes in Shah et al.'s paper are truly "civic," rather than political – i.e., Shah, et al. do not look at voting. Rather, they look at more localized community engagement, like going to meetings and volunteering for nonprofits.

Fast-forwarding seven years to Gil de Zúñiga, Jung, and Valenzuela (2012) ("GZJV") finds a paper that takes voting into account as one of many civic participation outcomes. GZJV considers the effect of getting news from social media on four outcomes: social capital, civic participation, offline political participation, and online political participation. Social capital is defined as "resources that can be accessed or mobilized through ties in [one's social] network" (Lin 2008, 51). Online political participation in GZJV (2012) includes making donations, emailing politicians, and subscribing to a political listserv (324). The remaining two outcomes split the civic behaviors used as outcomes in Shah et al. (2005) above: GZJV's (2012) "civic participation" outcome includes raising charitable moneys, attending neighborhood meetings, and being a conscious consumer, and their "offline political participation" outcome includes voting, putting up signs, and attending a rally (324).

Informing the methodology of this thesis, GZJV (2012) perform four sets of hierarchical regression analyses, one for each set of outcomes. They find that

> after controlling for demographic variables, traditional media use offline and online, political constructs (knowledge and efficacy), and frequency and size of political discussion networks, seeking information via social network sites is a positive and significant predictor of people's social capital and civic and political participatory behaviors, online and offline. (319)

Specifically, they find that getting news from social media was a statistically significant predictor of offline political participation at the $p < 0.001$ level. Additionally, GZJV (2012) find that age, education, and getting news from non-social media sources are all also statistically significant predictors of offline political participation.

Gil de Zúñiga was an author on another paper that bears heavily on this thesis; Jung, Kim, and Gil de Zúñiga (2011) found that political knowledge acts as a strong mediator of news media use's effect on offline political participation. However, they found that news exposure has only a limited direct effect on political knowledge (almost all of which is mediated). The authors conclude that simple exposure to information may not lead to much political learning, and that more intense processing of information may be required for knowledge to be gained. That fact notwithstanding, the effect was still non-trivial, which supports the focus of this thesis.

Jung, Kim, and Gil de Zúñiga (2011) offer a few additional helpful findings for this thesis. They confirmed the causal order of news use → political knowledge → voting, by running their model with each possible ordering of variables (421). They also hypothesized that "attitude strength" might be a confound that was not included in their analysis (424), which informs the use of party ID and ideology as confounds in this thesis.

Lastly, the measure used for political knowledge by Jung, Kim, and Gil de Zúñiga (2011) dealt with only "general" political knowledge – who is the current Speaker of the House, who is the current British Prime Minister, which state is Sarah Palin the governor of, and so on (418). In light of the discussion in the Political Knowledge section above, this measure may be problematic for several reasons. In their conclusions, the authors note that other types of knowledge should also be investigated, including candidates' issue stance knowledge, which "may be more relevant than general political knowledge in an election context" (425). As will be seen in Chapter Three, this thesis uses two measures of knowledge in analysis: "General" and "Personal." While both types ask respondents to identify the party of something (a legislative chamber or an elected official, respectively), knowledge of one's own representatives in government ("PersonalKnowledge") represents a slightly different dimension of political knowledge than the general identification questions used in Jung, Kim, and Gil de Zúñiga (2011). It may also be less subject to the widely-reported on gender bias in national and institutional rule-based knowledge questions, which are the type used in Jung, Kim, and Gil de Zúñiga (2011). Additionally, Jung, Kim, and Gil de Zúñiga (2011) calculate their knowledge measure as a simple sum of correct answers (418), in contrast to this thesis, which calculates knowledge as the proportion of correct answers out of attempted answers.

One variable that was used in Jung, Kim, and Gil de Zúñiga (2011) that is not in this thesis is internal political efficacy. The authors found that, in addition to political knowledge, internal political efficacy was found to be a strong mediator of the relationship between news use and political participation. In other words, based on the authors' findings, getting news from social media makes people feel more capable of influencing the political process, which leads them to do so (422). The CCES contains no variables that could be used to approximate political efficacy, so unfortunately, it was not considered in analysis.

Finally, another paper co-authored by Homero Gil de Zúñiga investigates the distinction between news consumption and active engagement. Weeks, Ardèvol-Abreu, & Gil de Zúñiga (2017) use the terms "opinion leaders" and "prosumers" to refer to people in their study who actively commented on, forwarded, or created news content on social media. The study focused on those people's self-perceptions and whether seeing oneself as an "opinion leader" makes one more likely to try to influence others' politics. They found that people who are more actively engaged with politics on social media *do* have outsized effects on other people's political attitudes and behavior, which grants policy relevance to RQ2 (15). That influence was corroborated by Turcotte et al. (2015), which found that people trusted a news story's source more when they believed the story had been posted by one of their real-life Facebook friends. The effect was even greater when the participant perceived the friend as an opinion leader.

In summary, the existing literature that has been reviewed here grants legitimacy to both main research questions posed at the end of Chapter One. The literature also supports the research methods that will be laid out in the next chapter, particularly by establishing a set of control variables that should be included in regression analysis. For clarity, I have summarized this chapter's findings regarding control variables in **Table 2** below, which indicates which of the three main variables (social media use, political knowledge, and voting) each potential confound is related to, and which sources in the literature provide evidence of that relationship.

In addition to giving legitimacy to the methods of this thesis in those parts held common, some of the literature cited in this chapter shows the novelty of the methods of this thesis. First and foremost, no paper has looked at social media news use, political knowledge, and voting in the specific context of the 2016 election. Given the nature of that election, there is ample reason to believe that people may have had a different relationship to social media than in previous political contexts. Additionally, the knowledge measure used in this thesis differs from that used in previous work and *may* be less gender-biased than knowledge measures used in previous work relating social media news use, knowledge, and voting.

Table 2: *Summary of Potential Confounding Variables*

Characteristic	***Using Social Media***	***Political Knowledge***	***Likelihood of Voting***
Education	1,2,7	5,6,11	7,8,11
Income	1,10	5,11	8,10,11
Region of the US		5	
Political Attitudes	10	6,10,11	10,11
Age	2,3,7,10	5,6,11	7,8,10,11
Race	1,2	5,11	8,11
Gender	1,4	5,9,11	8,11
Marital Status			8
Other News Use	7	7	7
State of Residence		5	

1: Scott et al. (2017), 2: Shearer & Gottfried (2018), 3: Smith & Anderson (2018), 4: Moore & McElroy (2011), 5: Brennan (2016), 6: Pew (2018b), 7: Gil de Zúñiga, Jung, and Valenzuela (2012), 8: MIT Election Data & Science Lab, 9: Dolan (2011), 10: Weeks, Ardèvol-Abreu, & Gil de Zúñiga (2017), 11: Kenski & Stroud (2006)

Methods

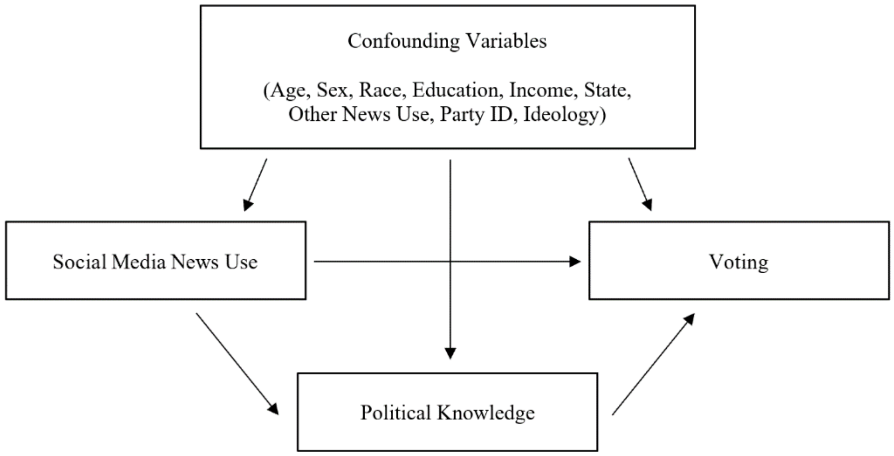
Drawing on the theoretical foundations established in the previous chapter, this chapter establishes statistical models that might provide answers to the two research questions of this thesis. The chapter begins by building a causal diagram to relate social media news use, political knowledge, and voting, using the potential confounding variables uncovered in the previous chapter. Then, an overview of the data source is provided, and some of its strengths and limitations are discussed. Next, variables are operationalized, which includes two exploratory factor analyses. Finally, the operationalized variables are combined with the causal diagrams to construct two sets of regression analyses.

1. Causal Diagram

Based on the predicted confounding variables identified in Chapter Two and summarized in **Table 2** above, a causal diagram relating social media news content, political knowledge, and likelihood of voting was constructed. That diagram is included as **Figure 1** below. Any characteristic that, based on the literature, was expected to influence two or more of the main variables was included as a potential confound in the graph.

Figure 1 is an example of a Directed Acyclic Graph or "DAG," and was made with DAGitty, a web-based tool for creating DAGs (Textor, et al. 2016). A DAG represents each variable in consideration as a node, with arrows connecting any two variables that are correlated. The direction of the arrow represents the direction of causality; so, for instance, the arrow from Age → Voting represents a direct causal path from age to voting, signifying that age has a direct effect on voting.

Figure 1: *The Mediated Relationship Between Social Media News Consumption and Voting*



It should be noted that there are many relationships between the confounding variables that have not been included in this diagram, for clarity. For example, age, sex, and race, are all strong predictors of both education and income, and education and income are themselves related as well. If one or more confounding variables did not have a direct effect on the outcome, i.e. did not have an arrow drawn from itself to voting, it would be necessary to chart each of these relationships in order to find and close the correct biasing pathways using Pearl's backdoor criterion (Pearl 1998, 255-259). However, each confounding variable has, based on the literature, a direct effect on both the treatment and the outcome. Therefore, all confounding variables that *can* be controlled for will *have* to be controlled for, making their inter-relationships, for charting purposes, moot.

Additionally, it should be noted that there are some predicted confounds in this relationship which cannot be controlled for – most importantly personality. Also omitted from this diagram are predictors which have not been theorized to be confounds – those variables from **Table 2** above which are theorized to affect only one of: treatment, mediator, and outcome.

2. The 2016 Cooperative Congressional Election Survey (CCES)

As mentioned previously, this thesis uses data from the 2016 Cooperative Congressional Election Survey (CCES). The CCES is unique among political surveys for several reasons. Chief among them for this thesis is that it asks respondents about their *active* online news-related activity – whether they read or watch political news content, share political information, put their own opinions out online, and so on. Other surveys, such as the American National Election Survey and ones conducted by the Pew Research Center, are superior on more traditional questions of media literacy (i.e. access to, frequency of, and attitude towards internet use), but do not include questions relating to active engagement with political news material.

The CCES is a national project, the collaborative product of sixty teams across the country. In 2016, each team surveyed roughly 1,000 people in two waves, pre-election (September 28 to November 7, 2016) and post-election (November 9 to December 14, 2016). Survey respondents were asked questions from three pools, which constitute three tiers of content: Common Content, Group Content, and Team Content. Teams and groups of teams have access to their unique content, while the Common Content is made publicly available through the Harvard Dataverse (Ansolabehere & Schaffner 2017). The Common Content data are the source for this thesis, and they include basic sample identifiers, including weights; demographic information, including ideology and voter registration status; pre-election questions; post-election questions, including vote-validation data; and contextual data, including candidate and incumbent information.

Survey responses were collected through YouGov, an online membership-based service that pairs respondents with researchers. Survey participation is voluntary and requires understanding of YouGov's privacy policy, which includes anonymization of any individual data provided to clients, such as the researchers behind the CCES (YouGov.com).

To gather the sample, YouGov used a matched random sample methodology for data collection. This method matches individuals in a sampling frame to a researcher-provided target population, using the same theoretical foundations as stratified sampling. First, both the target population and YouGov's member base are broken into the same mutually exclusive and exhaustive strata. For the 2016 CCES, over 150,000 YouGov members were pooled to construct these strata. Potential respondents were then recruited out of each YouGov stratum through simple random sampling until the proportion of each sample stratum within the total sample population matched its equivalent's frequency in the target population (Ansolabehere, Schaffner, & Luks 2017, 12). The sample for the 2016 CCES was later limited again to those respondents who completed both the pre- and post-election waves of questions, after which point weights were applied. The final product is a nationally-representative sample of 64,600 respondents.

YouGov's online sampling methodology is advantageous because of the large sample size it makes possible. This large sample size enables, among other things, state-level validation of the CCES sample, by comparing CCES two-party vote share predictions to actual state-level results. Graphs comparing the CCES estimate and actual two-party vote share in the presidential, senate, gubernatorial, attorneys general, and secretaries of state elections are included on pages seventeen to twenty-one of the 2016 CCES guidebook. They show that CCES estimates are a very good approximation of actual electoral outcomes, with over 95% confidence of a match between estimate and actual outcome in most cases (Ansolabehere, Schaffner, & Luks 2017, 17).

The main drawback of CCES data is YouGov's digital sample frame. All survey participants necessarily have at least some experience using digital technology and the internet, meaning the sample is likely to be biased towards people who use social media. The primary authors of the CCES acknowledge this limitation in CCES documentation when enumerating assumptions necessary to confirm the representativeness of the CCES sample (Ansolabehere, Schaffner, & Luks 2017, 12-13). Specifically, they reiterate the "common support" assumption, which holds that the range of values in variables used for sample matching must be plausibly the same for people included in the survey sample as for people not included in the sample frame. Because the survey is executed via computer, this excludes any variable measuring computer use from being a matching variable, because there cannot be anyone in the survey sample who never uses a computer. Because of that fact, it cannot be said with certainty that the survey sample and the target population (the US population, in this case) are the same on such metrics. However, there is nothing that can be done quantitatively to correct for this, and it can only be noted as a limitation.

2 A. Vote Validation in the CCES

A second problem spot – to avoid calling it a limitation – in the CCES is vote validation. Several research teams have independently identified discrepancies relating to validated vote calculations that can be performed using the CCES (Grimmer, et al. 2017; Agadjanian 2018a; Agadjanian 2018b; Ansolabehere & Hersh 2012). The problem is not limited to just the CCES, though; most surveys dealing with voting behavior consistently over-estimate voter turnout rates (Cuevas-Molina 2017, 1). Political scientists have postulated several explanations for this, including sample selection bias, a social-desirability effect, and a combination of the two (Ansolabehere & Hersh 2012, 439-441). Sample selection likely contributes at least some of the observed upward bias, as people who opt in to a survey about politics are more likely to care about and participate in politics. Social desirability also likely contributes to the observed bias;

there is a societal norm of voting, and people are reluctant to admit to not having done so, even in anonymous survey interviews. These may not only independently bias surveys' voter turnout metrics, but also interact, according to Bernstein, Chaha, and Montjoy (2001), who predict that "people who are under the most pressure to vote are the ones most likely to misrepresent their behavior when they fail to do so" (24). If people have self-selected into a political survey and therefore care more about politics, but for whatever reason(s) did or could not vote, they may be even more likely to make false claims than someone to whom politics is less important.

Whatever its causes, the effect is "non-trivial" in essentially all national surveys that deal with voting, and must be addressed (Cuevas-Molina 2017, 1). To do so, survey makers now use publicly-available voter registration records and privately-sold voter lists in order to validate people's reported voting behavior. Before releasing a data set, teams can match as many survey respondents to official voter records as possible, in order to confirm or contradict their self-reported voter registration status and voting history (Ansolabehere & Hersh 2012, 437). More than roughly ten years ago, researchers were limited in the extent to which they could do this, given irregularities in availability and quality of data. The CCES was the first project to perform fifty-state vote validation, in 2012 (439). Historically, even after performing vote validation, the CCES still overestimates voter turnout; Agadjanian (2018b) shows that the CCES overestimated national voter turnout in 2012 and 2014 by 6 and 14 points, respectively. These calculations were performed using CCES' validated turnout and voting-eligible for highest office turnout from the United States Election Project (Agadjanian 2018b; data available through McDonald n.d.).

The authors of the 2016 CCES suggest three ways to incorporate vote validation into analysis, in order to approve the accuracy of voter turnout (Ansolabehere, Schaffner, & Luks 2017, 126). One additional option not included in those three is to forgo vote validation altogether; whether someone says they voted is a decent indicator of political interest and using such a measure avoids the risk of creating false-negatives. However, doing so would also hamstring any causal claims that might be made; any findings would not relate to a person's actual likelihood of voting, but rather only to their likelihood of *saying* they voted.

Because of that limitation, it was decided that this thesis would use a validated vote measure. Therefore, a choice had to be made between the three validated vote measures outlined by the CCES authors. Those measures make use of information made available by a collaboration between the CCES team and Catalist, a political data firm. Researchers compared 2016 CCES respondents to a national database of registered voters prepared by Catalist, which resulted in 45,117 (69.84%) of those who completed both waves of the CCES survey being matched to a state's voter file. 19,483 (30.16%) were not.

The three vote validation measures suggested by the CCES authors are explained in more detail in **Table 3**. The first method, VV1, maintains the largest sample size, by making a non-voter out of anyone who was not affirmatively matched as a registered 2016 voter. This method relies on the accuracy of both government voter file data and CCES survey responses for matching, making its definition of non-voters the least rigorous. While this will unfortunately result in some false-negatives, the CCES authors point out that the most common reason that respondents couldn't be matched was that they were in fact not registered to vote; rates of self-reported non-registration/non-voting were "much higher" among those not matched to a voter file than among those matched (Ansolabehere, Schaffner, & Luks 2017, 126).

The second method, VV2, makes classifications only for respondents who could be matched to a voter registration; anyone not matched is left missing. This reduces the sample size significantly (dropping 30% of the original data set), though it has the benefit of reduced likelihood of false negatives.

Finally, while VV1 and VV2 employed in their classifications only the variables for being matched to a voter registration and for having cast a verified 2016 vote, VV3 also employs self-reported registration and voting behavior. Among respondents who could not be matched to a voter file, VV3 takes self-reported non-registered and/or non-voting people at their word. The CCES authors justify this assumption on the grounds that "self-reported non-voters are honest about their non-participation because there is no incentive to go *against* the democratic norm of participation" (126 – emphasis added).

Table 3: *Defining Three Validated Voter-Turnout Measures*

Measure	Voters	Non-Voters	Treated as Missing
VV1	*Matched to voter registration and vote confirmed as cast	*Matched to voter registration and vote confirmed as not-cast *Not matched to any voter registration	None
VV2	*Matched to voter registration and vote confirmed as cast	*Matched to voter registration and vote confirmed as not-cast	*Not matched to any voter registration
VV3	*Matched to voter registration and vote confirmed as cast	*Matched to voter registration and vote confirmed as not-cast *Not matched to any voter registration and self-reported not being registered to vote *Not matched to any voter registration and self-reported not casting a vote	*Not matched to any voter registration and self-reported not having cast a vote

SOURCE: Adapted from Ansolabehere, Schaffner, & Luks 2017, 126

Each validation method has its advantages and disadvantages, with none being clearly strictly-dominated or dominant on qualitative considerations. Empirics help to decide which measure to use in analysis. **Table 4** compares state-level CCES predicted turnout using each of the three measures to actual turnout. Comparisons are made to two different statistics for "actual turnout" from the United States Election Project: turnout among the voting-age population (VAP) and among the voting-eligible population (VEP).

Both the VAP and VEP comparisons are included, for separate purposes: Alexander Agadjanian (2018b) used the VEP in his comparisons, but VAP turnout is a better analogue for some estimates of CCES turnout, when respondents are not limited to eligible voters. Whether the CCES sample is more like the voting eligible or voting age population changes depending on the VV method used. When using VV1, the sample is more similar to the VAP, because non-matched respondents are considered simply non-voters, rather than excluded as missing; voting non-eligible people are still included as the sample, as non-voters, rather than excluded completely. For VV2, the sample is more like VEP, because the sample is limited only to matched registrants, meaning that no voting-non-eligible people can be in the sample. For VV3, neither VAP nor VEP is a great match, because VV3 uses includes some non-matched respondents in the sample, but not all.

Ultimately, precision about each VV measure's comparability to VEP versus VAP did not bear great significance, because as **Table 4** shows, VV1 is the closest estimate for *both* types of actual turnout, VAP and VEP. Among the 51 state-level turnout estimates (50 + DC), VV1 is off from VAP turnout by an average of less than one percentage point, far less than any other measure.

Table 4: *Selecting a Validated Voter-Turnout Measure*

Turnout Measure:	***VV1***	***VV2***	***VV3***	***VV1***	***VV2***	***VV3***
Compared to Turnout Among:	*Voting Age Population*			*Voting Eligible Population*		
Average Percent Error (%)	8.38	42.15	28.41	9.31	32.07	19.36
Average Bias (%-Pt. Diff)	0.91	22.98	15.41	-3.25	18.83	11.26
Average |Bias|	6.65	31.05	21.49	0.56	24.96	15.40

SOURCE: Actual Turnout Data from McDonald (n.d.): www.electproject.org/2016g

Agadjanian's (2018b) findings also appear to support the selection of VV1 as the voter-turnout measure of choice. Agadjanian's figures report that 2016 CCES state-level turnout estimates are off from VEP turnout by an average of -4 percentage points. This is quite close to the -3.25 average bias figure that was calculated for VV1 as compared to the same actual turnout measure (VEP). The two average bias estimates are not exactly equal, which is of some concern, but it may be due to a difference between the 2016-only CCES data set (used in this thesis' analysis) and the 2006-2016 CCES Cumulative File (used in Agadjanian's analysis).

As such, VV1 is the best measure of voter turnout to use in this analysis, because it minimizes state-level bias in turnout. It also is the best estimate of national voter turnout, calculated as a simple weighted mean of individuals in the data set. Nationwide turnout was 59.2%/54.7% (VEP/VAP) in 2016; VV1, VV2, and VV3 produce weighted means of 55.8%, 79.0%, and 71.0%, respectively. It should be noted that the analysis in this thesis is primarily at the individual level and not the state level, so the problem of state-level bias was, even at the start, not as pressing. Indeed, the CCES is currently one of the premier national data sets for studying individual-level political behavior (Grimmer, et al. 2017, 2). Nonetheless, it is best to seek the best measure possible, and state-level turnout will be included as a potential confound in analysis through the inclusion of state fixed-effects.

3. Modeling Social Media Behavior

The next step to prepare the CCES data for analysis was to devise measures for the two types of news-related social media behavior described in RQ1 and RQ2: consumption and active engagement. To create those measures, explanatory factor analysis (EFA) is performed on five variables from the CCES.

As introduced above, the CCES is unique in that it asks respondents above their active social media posting, not just their social media use. Also distinctive about these questions is that they ask the Yes/No question of whether respondents engaged in a given activity *in the last twenty-four hours,* rather than how frequently they may have done so over a given period of time, such as the last week. Framing questions this way should increase the accuracy of the measure. People are notoriously poor estimators, including when it comes to their own behavior (Brennan 2016). Asking respondents to remember only whether they did something at all over the last twenty-four hours of their lives, rather than summarize the frequency of doing so over the last week, or month, theoretically increases the odds that they provide an accurate answer.

All five questions used to create social media scores were asked in the pre-election wave. Respondents were first asked whether they had used social media in the past twenty-four hours. 19,286 respondents had not. The 45,314 respondents that answered in the affirmative were asked a follow-up question which asked whether they did each of the following on social media in the last twenty-four hours:

1. "Posted a story, photo, video or link about politics"
2. "Posted a comment about politics"
3. "Read a story or watched a video about politics"
4. "Followed a political event"
5. "Forwarded a story, photo, video or link about politics to friends"

Given the five variables, the total number of possible response patterns is 32. The relatively small size of this set of possible response patterns (Bartholomew, Knott, & Moustaki 2011, 84) allows it to be displayed as a frequency distribution (included in the Appendix as **Table 10**). Brief examination of this distribution allows some insights. The most frequent response pattern, at 22.69% of social media users, is a respondent having read or watched a story about politics in the past twenty-four hours but not having engaged in any of the remaining four activities. The next two most frequent patterns are markedly less than the most frequent response. These second and third-most frequent response patterns are also the two extremes: having engaged in none of the activities in the past twenty-four hours and having engaged in all of them. Having engaged in none of the activities is roughly 5 percentage points less frequent than the most frequent pattern, coming in at 17.45% of social media users, and the jump from none to all is even more drastic, with only 10% of social media users engaging in all six activities (a difference of roughly 7.5 percentage points).

With an understanding of the pattern of people's responses, the next step was to attempt to analytically differentiate between the two types of social media news use, consumption and active engagement. Factor analysis allows that difference to be teased out specifically within the analytical sample of this thesis.

Factor analysis attempts to uncover and interpret patterns and relationships within data by identifying a number of "factors," unobserved traits, that predict "measures," observed behavior. Factors are also referred to as "latent" and "unobserved" traits. Factor analysis can be performed on a variety of combinations of variable types of factors and measures; social scientific scholars increasingly using factor analysis to isolate continuous factors from dichotomous measures, a fundamental form of latent variable analysis.

EFA works by testing a range of hypotheses regarding the relationships between the observed variables, narrowing the set of hypothetical relationships to find which variables tend to "go together" (Yong & Pearce 2013, 80). The final output of EFA is the smallest set of factors that accounts for the highest amount of correlation amongst the observed variables (80). Ideally, each variable will be highly correlated with one factor and not strongly correlated with any other. The correlation of each variable with each factor is called a "loading," and the table of each variable's loading onto each factor is called a "pattern matrix." Hair et al. (1998) provide thresholds for practical significance of factor loadings: ≥ 0.3 = "minimal," ≥ 0.4 = "more important", and ≥ 0.5 = "practically significant" (111). The "communality" of a variable in a factor analysis is the sum of its squared loadings, and can be interpreted as the proportion of total variation in that variable that is explained by the factors.

The factor analysis below was performed in Stata/IC 15.1, as was all other analysis. Stata's factor command was used to perform EFA, using the iterated principal-factor option (IPF) based on the recommendation of Habing (2003). Whereas other methods for fitting factor analysis models use the same initial estimate of each variable's communality throughout each successive attempted model fit, the iterated principal factor method recalculates the communalities with each new fitted model (Habing 2003, 5). Habing (2003) recommends IPF so long as the final model has neither impossible communalities (greater than one) nor negative error variances (5). As is seen below, neither is the case in this model.

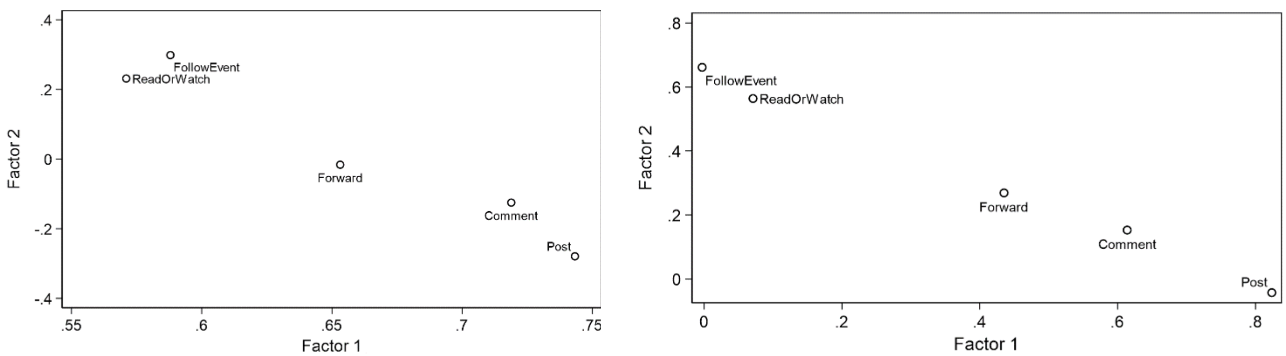
An initial EFA of the five social media behavior variables yields a four-factor model (pattern matrix included in the Appendix as **Table 11**). This is not much of a reduction in variables. Importantly, the third and fourth factors are very lightly loaded; most loadings onto them are less than 0.1, and the greatest in terms of absolute value is 0.14. According to Hair et al.'s (1998) thresholds, these loadings are of less than even "minimal" practical significance. On those grounds, further analysis is limited to only the first two factors.

After limiting the EFA to only two factors, Stata returns an improved pattern matrix with more-significant loadings (included in the Appendix as **Table 12**). Posting, commenting, forwarding, and following all have practically significant loadings onto at least one factor, and reading/watching is slightly more than minimally significant. However, another potential problem remains: multiple variables load relatively-significantly onto more than one factor. This will make interpretation of the factors more difficult, and according to Habing (2003), should be avoided if possible.

To correct this, the factors can be rotated. Dr. Maike Rahn of TheAnalysisFactor.com provides a very intuitive explanation of factor rotation. As Dr. Rahn describes it, the two factors are axes along which each variable is positioned. When Stata conducts the initial EFA, it isolates the first factor (axis), holds that constant, then finds the second. Once these two factors (axes) have been found, however, they do not always result in the best possible fit of the data. Rotating the axes can make them fit the variables better. A side-effect of rotation is that it ideally aligns each variable primarily with only one factor, thereby making the factors more easily interpretable.

There are two main families of rotation techniques: orthogonal and oblique. Orthogonal rotations preserve the original EFA's assumption that the factors are not correlated. Continuing to use the analogy of axes, an orthogonal rotation preserves the ninety-degree angle between the axes. An oblique rotation relaxes that assumption and allows the factors to be correlated. In the axis analogy, the angle between newly-rotated axes is allowed to no longer be exactly ninety degrees. As seen in **Figure 2** below, rotating the factors does improve the overall fit to the data, making the relationship between variables, plotted on the two factors, more linear.

Figure 2: *Social Media EFA Loadings, Before and After Rotation*



In this case, an oblique rotation was used for the social media behavior factor analysis. Theorizing that the two factors are baseline social media news consumption and active social media posting (akin to Scott et al.'s (2017) "frequency" and "engagement"), it should be expected that the factors are somewhat correlated. The results of this final factor analysis are presented in **Table 5** below. After rotating the factors and allowing them to be correlated, most variables settle nicely onto one of the two factors. As a result, the interpretation of the factors as propensity for active social media posting and propensity for baseline social media use, respectively, becomes clear. Posting and commenting both load highly and primarily onto the first factor, seemingly because each involves the active moving-around and/or creation of political content. Reading/watching content and following an event load heavily and primarily onto the second factor, likely because they are more passive activities. Interestingly, forwarding content loads more heavily onto the first factor, but also loads onto the second factor (with a loading four times greater than any other variable's non-dominant loading). This is makes sense; forwarding content would require first consuming it.

Table 5: *Social Media Behavior EFA Results*

Variable	Active Engagement	Consumption	*Communality*
Post Content	0.8236	-	0.6303
Post Comment	0.6138	-	0.5324
Forward Content	0.4351	0.2691	0.4269
Read/Watch Content	-	0.5640	0.3796
Follow Event	-	0.6614	0.4349

*Factor loadings ≤ 0.1 omitted for clarity. Version including omitted loadings available in the Appendix as Table 13.
**Iterated Principal Factors Method, Oblique Promax Rotation, 2 Factor Limit

With the two factors now identified, a score could be created for each. Two main approaches were considered: (weighted) sum scores and regression scores. The benefit of sum scores is interpretation; in regression analysis, an increase of one unit for a sum score simply means engaging in one more activity, while an increase of one unit for a regression score is much less clear to interpret. Sum scores, however, give each variable equal weight, when in reality, such may not be the case (DiStefano, Zhu, & Mîndrilă 2009, 8).

In the case of the Consumption factor, giving equal weight to the two variables which load most heavily would not make an irreconcilable assumption. Although following an event does load more strongly onto the factor than does reading/watching content, the difference is not so large as to suggest a very large difference in importance. In contrast, it is reasonable to suspect that the variables going into the active engagement score should not be weighted equally; as seen in Table 5, the variables load significantly differently onto that factor, suggesting that certain variables (e.g. posting content) are significantly more related to the underlying factor than others (e.g. forwarding content).

One method to account for this unequal importance of variables is to create weighted sum scores, by using variables' loadings as weights. This gives the most important variables the most weight on the resulting score, though the benefit of doing so is premised on the loadings representing true differences in the importance of the variables. When doing so, the fact should be considered that it is always possible that the differences implied by the loadings are not factual, but rather are simply the effects of choices made during the EFA process, such as rotation (DiStefano, Zhu, & Mîndrilă 2009, 8).

A second method to account for the unequal importance of variables is to use OLS regression to predict each respondent's location on each factor. This is the default prediction method in many statistical packages, including Stata. It differs from a weighted sum in that it takes into account not only the relationship between the variables and the factors (through the loadings), but also among the variables themselves, and between the factors, if they have been allowed to be correlated (which, in this case, they have) (DiStefano, Zhu, & Mîndrilă 2009, 4). Scores computed using the regression method are standardized, with a mean of zero and a standard deviation equal to the squared multiple correlation between the factor and its constituent variables (4).

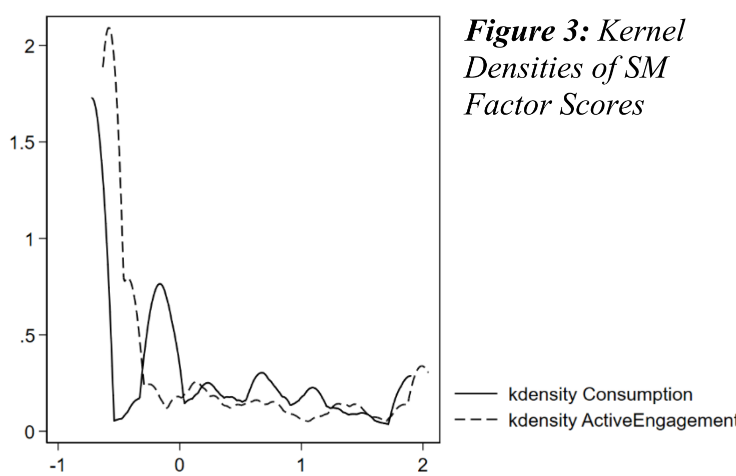


Figure 3: *Kernel Densities of SM Factor Scores*

Regression was ultimately the method used to create scores for Consumption and Active Engagement. The justification for this choice is twofold: first, regression best accommodates for the split loading of forwarding content, seen in **Table 5** above. Secondly, the correlation between the two factors is most easily accounted for through the regression method. The distributions of the two scores created through that method are plotted in **Figure 3** above; as that figure shows, both scores are heavily skewed to the right, with many respondents falling at the scores' minimum values. The scores generally appear to follow a power law distribution, matching previous findings regarding the distribution of social media use in Java, Song, Finin & Tseng (2007) and Leskovec et al. (2007).

4. Modeling Political Knowledge

In addition to a model of respondents' social media news use, subsequent analysis requires a model of respondents' political knowledge. Again, EFA was used to separate questions into two groupings, using many of the same assumptions and techniques as the social media behavior EFA above. However, final scores were manually created, rather than through the regression method as above, because doing so both improves interpretation and allows two sets of scores to be created, differing in their treatment of "Not Sure" responses.

The political knowledge EFA is performed on six measures: respondents' ability to correctly identify the parties of the current Congressional majority, the current Senate majority, their U.S. Representative, their two U.S. Senators, and their Governor. A "correct response" was coded by comparing the respondent's guess to contextual variables containing the actual party of the chamber/officeholder in question. That process was repeated for each of the six variables, to create six binary variables, indicating accuracy of response to each question. Given these six items, the total number of possible response patterns is 64. This again relatively small set of possible response patterns is displayed as a frequency distribution in the Appendix, **Table 14**.

Again, brief examination of the frequency distribution provides some insights. 35% of the sample answered all six questions correctly. Another 12% failed to answer a single correctly. Of the remaining 53%, a number of combinations were relatively more common, each appearing between four and five percent of the time in the total sample: correctly identifying individuals but not chambers, correctly identifying all but the respondent's Representative, and correctly identifying all but the party in control of the Senate. In the next tier, occurring roughly 2.5% of the time in the total sample, respondents correctly identified only their governor or correctly identified all but the party in control of the House.

It is worthwhile to compare the performance of CCES respondents on political identification measures to those cited in Chapter Two above. As a reminder, a 2014 APPC survey found that 38% of Americans could correctly identify the party in control of the U.S. House, and the same percent could do so for the Senate. In the CCES data set, the success rate is considerably higher: 62% of respondents correctly identified the Republican majority in the House and 57% did so for the Senate.

This difference may be due to two factors, both related to timing. The APPC survey was conduct July 8-14, 2014. In contrast, the CCES political knowledge questions were asked in the pre-election wave of the survey that was in the field from September 28 to November 7, 2016. As such, the CCES questions were asked anywhere from two to four months closer to the Election Day of the given year than were the APPC questions. Secondly, the 2016 election included the presidential race, while the 2014 election did not, and presidential elections are widely known to garner higher visibility and generate increased interest in politics at all levels. Because of these effects, politics were likely relatively-higher salience around the time of the CCES, as compared to the time of the APPC survey, explaining some of the difference.

The second political knowledge statistic cited in Chapter Two that can be compared to CCES estimates is Americans' knowledge of the party of their Representative in Congress. A 2017 survey found that 56% of Americans claimed to know their personal Representative's party (Freiling 2017). CCES respondents accurately identified the party of their Congressperson 60.1%

of the time – fairly close to the statistic from Freiling (2017). Still, there are immediate differences between the two calculations; the Freiling figure represents respondents' *claim* of knowledge, whereas the CCES figure represents respondents' actual knowledge. Delving deeply into the differences between the two is not the purpose of this thesis. It is simply interesting and, to some extent, affirming that the figures are relatively similar.

Returning to the full set of six CCES questions, as with social media behavior, a way to accurately reduce the six items into a more workable set of measures is needed. Factor analysis allows the various types of party identification to distinguished from one another. As for social media use, the iterated principal factors method is used. IPF is in this case again acceptable because neither impossible communalities nor negative error variances are present. Running an initial EFA with no limit on the number of factors returns a set of five factors (see Appendix, **Table 15**). Again, as was the case in the social media use EFA, only two of the factors had variables load onto them by more than 0.15. Therefore, the EFA was limited to only two factors. The two-factor limited EFA-IPF returned a much-improved pattern matrix in which most loadings are roughly 0.4 or greater (see Appendix, **Table 16**). However, two variables again loaded onto more than one of the factors, with non-dominant loadings just shy of 0.4. Therefore, the factors were again rotated, using an oblique rotation, because types of knowledge may be correlated. Based on the rotated loadings, included in **Table 6** below, two clusters appear: one which represents knowledge of one's own personal representation in government and one representing knowledge of broader balances of political power. Conceptually, it makes sense for those two factors to be distinct; someone can know the general state of things in Congress, but not pay much attention to their exact own representatives, and vice-versa.

Table 6: *Political Knowledge EFA Results*

Variable	PersonalKnowledge	GeneralKnowledge	*Communality*
House Majority	-	0.7650	0.6498
Senate Majority	-	0.7695	0.6280
Governor	0.6753	-	0.5178
Representative	0.6642	-	0.4866
Senator 1	0.7394	-	0.5823
Senator 2	0.6789	-	0.5260

*Factor loadings ≤ 0.1 omitted for clarity. Version including omitted loadings available in the Appendix as Table 17.
**Iterated Principal Factors Method, Oblique Promax Rotation, 2 Factor Limit

With variables now separated by factor, scores could be created. In contrast to the social media scores above, a simple sum method was used to created scores for political knowledge. This is similar to the way index scores were created in Jung, Kim, and Gil de Zúñiga (2011). In contrast to the social media EFA above, variables per knowledge factor here load roughly equally; Senator 1 loads slightly more than the other three variables that load onto personal knowledge, but not to an extent that causes concern.

Two sets of scores were created for the knowledge factors, to examine the effects of taking into account Fortin-Rittberger's (2016) findings regarding the exclusion of "Not Sure" responses. Quick analysis of the six questions used provides strong evidence for the gender guessing gap; after applying survey weights, the group of respondents who responded "Not Sure" on at least one political knowledge question was 40% male versus 60% female.

The first set of scores created were the "standard" scores, which simply calculate the proportion of questions per factor that the respondent correctly answered. For these scores, when a respondent answered "Not Sure," it was coded as an incorrect answer. The second set of scores codes such "Not Sure" responses as missing values, rather than incorrect. These are the scores that attempt to incorporate Fortin-Rittberger's (2016) findings, and they represent the number of questions a respondent answered correctly *when they actually attempted to answer the question*.

All four scores have ranges [0,1]. Their distributions are discussed in more detail in the descriptive analytics section of Chapter Four. Hereafter, when the two sets of scores are referenced, the standard scores may also be referred to as the "Incorrect" scores, and the Fortin-Rittberger-esque scores may be referred to as the "Missing" scores – referencing their respective treatments of "Not Sure" responses.

Although calculating these scores as proportions means that interpretation is not as neat as a simple sum, the added benefit of accounting for some of the gender bias in political knowledge questions makes the trade-off by far worth it. Additionally, interpretation of these proportional scores is still easier than regression score output, which would have been the default choice.

5. Other Selected Variables for Analysis

The final step in data preparation was to recode the available variables for the confounding factors identified in Chapter Two. Age was created by subtracting the respondent's reported birth year from 2016. Female (0 male, 1 female) was created from the native CCES variable for gender. Income was created from the native variable faminc, preserving faminc's categories up to $149,999, but collapsing higher-income categories into a single "$150,000 or more" category in order to increase bin sample size. Similarly, the educ variable was bottom-coded, combining "No HS" and "HS Grad" into a single "HS or Less" category, to increase bin sample size – the "No HS" category contained less than 500 respondents in the original data.

Race was left mostly untouched, besides changing the order of values so that “Middle Eastern” came before “Mixed” and “Other,” rather than after – purely for display purposes. The respondent’s state of residence was accounted for with the native CCES variable inputstate.

For ideology and party ID, the simplest-possible variables were used. In other words, no degrees of strength were included, such as “Strong Democrat” vs. “Lean Democrat” – respondents were simply Democrats. For party ID, the possible values were Democrat, Republican, Independent, Other, and Not Sure. For ideology, the possible values were Liberal, Moderate, Conservative, and Not Sure.

The reason for using these simpler measures that do not account for strength of attachment is two-fold: first, after looking into the CCES’ party ID variables a good deal, it seems as though there is some erroneous self-sorting by respondents that was not intended by the survey designers. The potential responses for the most-specific party ID variable were: Strong Democrat, Not Very Strong Democrat, Lean Democrat, Independent, Lean Republican, Not Very Strong Republican, Strong Republican, and Not Sure. When that variable was used to predict voting, an unexpected pattern emerged. Rather than voting increasing in likelihood as party ID increased in strength (from Independent to Lean to Not Very Strong to Strong), it instead increased from Independent to Lean, but dipped at Not Very Strong, then increased again at Strong. Based on that pattern, it is suspected that respondents interpreted the “Lean” and “Not Very Strong” labels in an unusual way, perhaps reversing the order of magnitude in their heads, despite the order of listing.

For that reason, simple identifications, without strengths, were used. The “center” category was used as the reference category for both variables in regression – for ideology, Moderates, and for party ID, Independents. Additionally, the interaction between these two variables was also included in all regressions, since it is reasonable to expect that people among the same party ID group might behave differently based on their ideology. This is supported by several commonly-identified trends, in which people have complicated relationships with party labels. It has been frequently reported that partisan independents who do not claim a partisan ID often still essentially act like partisans (Hawkins & Nosek 2012, Smith 2016). Additionally, as of 2008, the American public was trending towards conservative ideologies, but party IDs were not shifting as rapidly (Norrander & Wilcox 2008). The interaction between these two variables in the 2016 election is explored in the Descriptive Analysis section of Chapter Four below, which shows that there are some party IDs within which there is distinct variation in ideology.

Lastly, the control variable for having gotten news from other (non-social media sources) was constructed from three CCES variables: CC16_300_2, CC16_300_3, and CC16_300_4. These variables record whether respondents "watched TV news," "read a newspaper in print or online," and "listened to a radio news program or talk radio" in the past 24 hours, respectively. OtherNewsUse was set to 1 if the respondent had done any one or more of these three items; it was set to 0 if they had done none.

6. Regression Models

With the causal diagram, social media and knowledge scores, and other recoded variables in hand, regression models could now be constructed. The first set of regressions build on the causal diagram created in **Figure 1**, and seek support for **H1:** consuming political news through social media will be positively related to voting, and **H2:** the effect of social media news use on voting will be partially mediated by an increase in political knowledge. In that set, I include six separate models, distinguished by their inclusion/exclusion of state fixed effects, the presence of mediating variables, and the type (coding) of those mediating variables. Those models are summarized in **Table 7** below. As that table shows, there is one model for each possible combination of the three defining characteristics; this will allow comparison of models where only one key aspect has been changed.

Table 7: *Descriptions of Voting Regression Models*

Model Number	**1**	**2**	**3**	**4**	**5**	**6**
State Fixed Effects		✓			✓	✓
Mediation Variables			✓	✓	✓	✓
Coding of Mediation Variables (Treatment of "Not Sure" Responses)	N/A	N/A	Missing	Incorrect	Missing	Incorrect

The first regression in the set is a multivariate regression of the outcome (voted) on the treatment (social media use), with controls, but without the knowledge variables. Logit regression is used due to the binary nature of the outcome variable. Both Consumption and ActiveEngagement, as well as an interaction term between the two, are included, because the two scores were allowed to be correlated when created using the regression method, meaning we therefore might expect that their effects are commingled. The variables for party ID and ideology are also interacted with each other. Model (2) is the same as model (1) with the simple addition of state fixed effects ($+ \beta_{13}{}^{*}State_{i}$).

$$\begin{aligned}\text{logit}(voted_{i}) = {} & \beta_0 + \beta_1{}^{*}Consumption_{i} + \beta_2{}^{*}ActiveEngagement_{i} \\ & + \beta_3{}^{*}Consumption_{i}{}^{*}ActiveEngagement_{i} + \beta_4{}^{*}Age_{i} + \beta_5{}^{*}Female_{i} + \beta_6{}^{*}Race_{i} \\ & + \beta_7{}^{*}Educ_{i} + \beta_8{}^{*}Income_{i} + \beta_9{}^{*}PID_{i} + \beta_{10}{}^{*}Ideology_{i} + \beta_{11}{}^{*}PID_{i}{}^{*}Ideology_{i} \\ & + \beta_{12}{}^{*}OtherNewsUse_{i} + \varepsilon_{i} \qquad (1)\end{aligned}$$

Next, I add in the score variables for general and personal political knowledge to act (theoretically) as mediators. There are four such regressions – models (3) through (6). The first two include the mediating variables but do not include state fixed effects. Such an equation is included below.

$$\begin{aligned}\text{logit}(voted_i) = {} & \beta_0 + \beta_1{}^*Consumption_i + \beta_2{}^*ActiveEngagement_i \\ & + \beta_3{}^*Consumption_i{}^*ActiveEngagement_i + \beta_4{}^*GeneralKnowledge_i \\ & + \beta_5{}^*PersonalKnowledge_i + \beta_6{}^*Age_i + \beta_7{}^*Female_i + \beta_8{}^*Race_i + \beta_9{}^*Educ_i \\ & + \beta_{10}{}^*Income_i + \beta_{11}{}^*PID_i + \beta_{12}{}^*Ideology_i + \beta_{13}{}^*PID_i{}^*Ideology_i \\ & + \beta_{14}{}^*OtherNewsUse_i + \varepsilon_i \qquad (3/4)\end{aligned}$$

However, that equation represents two separate regressions, models (3) and (4). Those models are distinguished from each other by the coding of GeneralKnowledge and PersonalKnowledge; model (3) uses versions of those variables in which “Not Sure” responses were coded missing, thereby dropping from analysis people who responded “Not Sure” on both General or all four Personal knowledge questions. This type of model has been referred to throughout this thesis has a “Missing” model. Model (4), then, uses knowledge measures in which “Not Sure” reponses were coded as incorrect guesses, thereby keeping more respondents in the analytical sample. The expoentiated coefficients on Consumption in (1) and (2) will be used in identifying any mediation resulting from the addition of the knowledge variables Models (5) and (6), then, are the same as models (3) and (4), respectively, with the exception of the addition of state fixed effects (+ $\beta_{15}{}^*State_i$).

Results

In this chapter, I present the results of the models developed in Chapters Two to Three. Before presenting the results of the regression analyses, I first review the operationalized variables from Sections 3.2 through 3.5, to provide an understanding of the data.

1. Descriptive Analysis

Descriptive statistics for all variables used in analysis are included in **Table 8** below. That table is split, to provide summary information for the variables in the differing contexts of the various models outlined in Section 3.6 above.

Table 8: *Weighted Descriptive Statistics of Key Variables*

Knowledge Score Type (Treatment of "Not Sure" Responses")	**Missing**		**Incorrect**		***Common to Both***	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Min**	**Max**
Consumption	.0770	.8608	.0034	.8222	-.7226	1.8994
ActiveEngagement	.0663	.9151	-.0021	.8757	-.6320	2.0459
GeneralKnowledge	.8037	.3355	.5950	.4426	0	1
PersonalKnowledge	.9015	.2144	.6769	.3679	0	1
Age	48.7264	17.1670	46.7614	17.1736	18	95
Female	.4613	.4985	.5138	.4998	0	1
Educ	3.4604	1.4222	3.3029	1.3857	2	6
Race	1.5810	1.3271	1.6070	1.3348	1	8
Income	6.4262	3.1503	6.0341	3.1564	1	12
PID	1.9858	.9360	2.0924	1.0704	1	5
Ideology	2.1768	.8318	2.2604	.8221	1	4
OtherNewsUse	.9118	.2836	.8638	.3430	0	1
Voted	.6115	.4874	.5590	.4965	0	1
InputState	28.2899	15.8287	28.3440	15.7600	1	56
Observations	35,120		46,359		N/A	

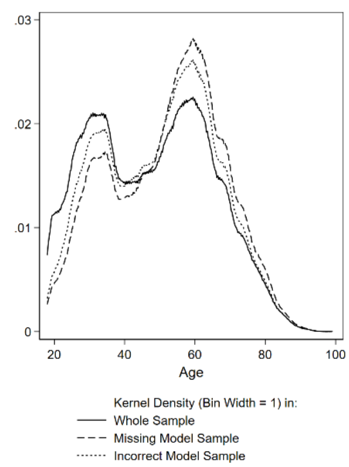
As that table shows, the distributions of certain variables are similar across model types, while others differ. Descriptive statistics are not split out by presence of inputstate in the model, as all observations have a value for that variable, meaning that samples with and without that variable in the model have identical samples. Additionally, the minimum and maximum values of all variables are the same across model types; therefore, only the variable means and standard deviations are broken out by model.

Based on **Table 8**, it is clear that the coding of "Not Sure" responses to knowledge questions impacts the sample in meaningful ways. First, the sample is over five percentage points less female when "Not Sure" responses are coded as Missing. The 22% of respondents who responded "Not Sure" to both GeneralKnowledge questions were two times as likely to be women than men. The 11.4% of respondents who responded "Not Sure" or "Never Heard of the Person" to all four PersonalKnowledge questions were 75% more likely to be men than women. Therefore, it follows that dropping those people would make the sample relatively more male.

Out of the remaining people – respondents who attempted to answer at least one question on each knowledge score – men were still more likely to have higher accuracy on the questions they attempted. However, the gap was *smaller* for the PersonalKnowledge score than for the GeneralKnowledge score. Men were 30% more likely to have 100% accuracy on the GeneralKnowledge questions (40.4% of men scored 1.0 on GeneralKnowledge vs. 31.0% of women). However, men were only 25% more likely to have 100% accuracy on the PersonalKnowledge questions (43.4% M vs. 34.6% F). The reduced gender disparity on PersonalKnowledge as compared to GeneralKnowledge may be evidence to support the claim made in Chapter Two that localizing/personalizing political knowledge questions is one way to reduce gender bias, as found in Miller (2018).

Additionally, voter turnout is roughly five percentage points higher in the "Missing" sample, suggesting that the people who are dropped as a result of coding "Not Sure's" as missing were majority non-voters. There is also a roughly two-year difference in mean age between the samples as well, indicating that the people dropped from the "Missing" sample are generally young. Still, the distribution of age in both samples is similar to the "true" distribution as reported by the Census Bureau. The two modes, around 30 and 60 years old, match the documented pattern in the United States (Rogers 2016); the Baby Boomer generation is in their late fifties to mid-sixties, and their children's generation is in their late twenties to mid-thirties. Both of these groups were larger than the generations before and after them (the Baby Boomer's children's generation *because* there were so many Baby Boomers giving birth), which led to the "lumpy" appearance of the age distribution in the US today.

Figure 4: *Age Distributions by Model Sample*



The treatment and mediating variables' characteristics also differ across the two samples. Those scores' distributions are plotted in **Figure 5**, **Figure 6**, **Figure 7**, and **Figure 8**. Both social media scores' means are actually closer to zero in the "Incorrect" (larger) sample – given that their means were set to zero in the process of creation and the Incorrect sample has over 10,000 more observations than the Missing sample, this is relatively unsurprising. However, the scores' distributions are otherwise very similar, as seen in the two figures below. The distributions are highly right-skewed, with most people occupying low score regions.

Figure 5: *Distribution of Social Media Consumption by Model Sample*

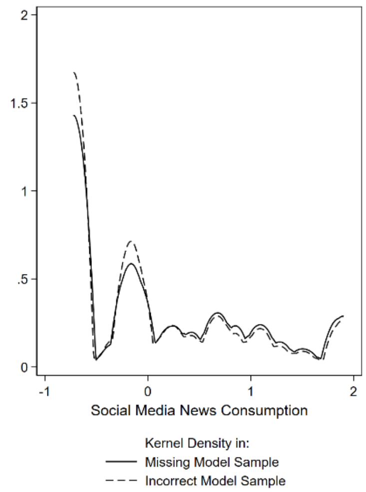
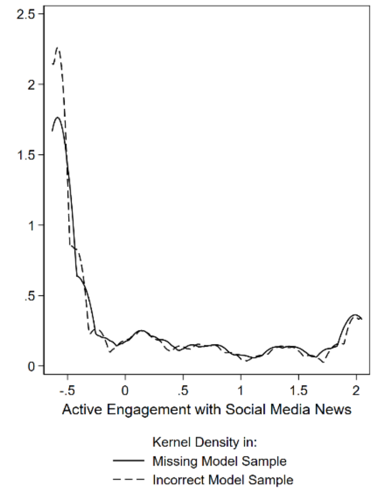


Figure 6*: Distribution of Social Media Engagement by Model Sample*



In the knowledge scores, there are over 20 percentage point differences in means across model samples (20 and 23, specifically). These are significant, but again unsurprising – these are the variables distinguishing the two samples, and coding over 10,000 "Not Sure" responses as incorrect answers is undoubtedly going to shift the distribution of the knowledge variables, as seen in **Figure 7** and **Figure 8**. Still, even with those responses coded as zero's, people answered knowledge questions correctly more often than not.

Figure 7: *Distribution of Personal Knowledge Scores by Model Sample*

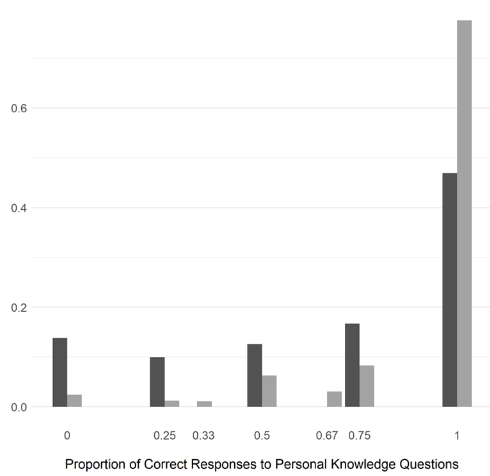
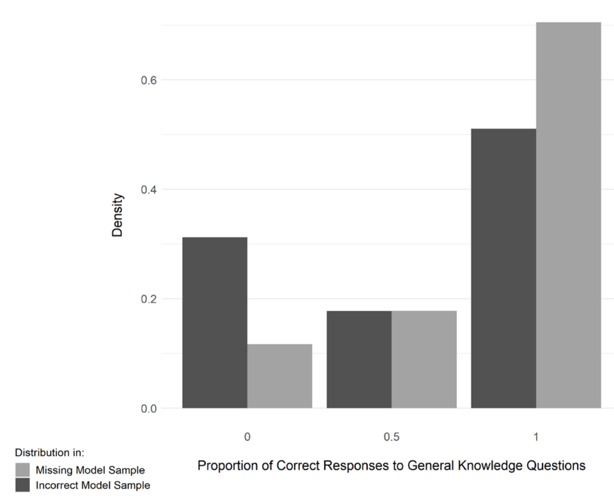
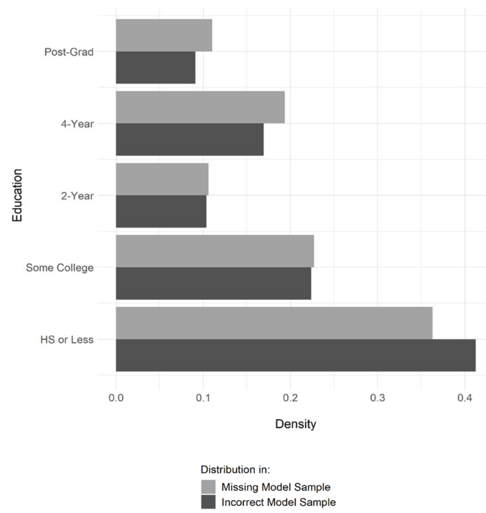


Figure 8: *Distribution of General Knowledge Scores by Model Sample*



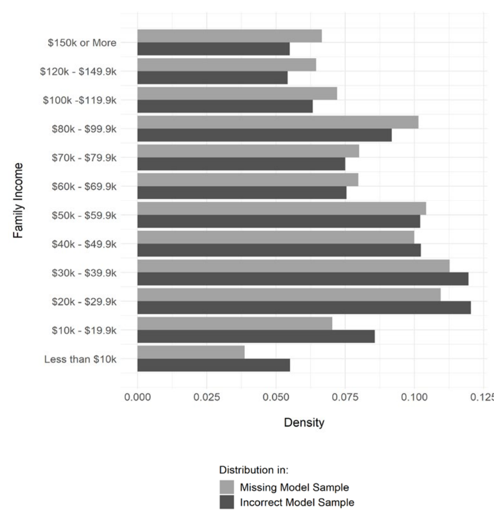
OtherNewsUse's weighted means of .91 and .86 indicates that most people in the survey sample got news in the 24 hours preceding their interview from at least one source that was not social media, whether it was TV news, the newspaper, or talk radio. This is relatively unsurprising.

Figure 9: *Distribution of Educational Attainment by Model Sample*



Among the remaining variables, there is little difference between the two samples. Race is generally appropriately distributed in the samples, with roughly 75% of respondents self-identifying as whole and roughly 12% as black, compared to 76.6% and 13.4% in Census estimates (Census QuickFacts). The survey data also approximate the reality of educational attainment in the US, with roughly three in ten Americans having a Bachelor's degree or above. The distribution of income also generally matches reality. The median family income level in both samples is the $50,000 - $59,999 bracket; the median household income as reported by the Census Bureau for the time period that includes 2016 was $57,652 (in 2017 dollars). More detailed distributions of income in the samples are plotted in **Figure 10**.

Figure 10: *Income Distribution in Analytical Sample*



Next, I explore the analytical sample by political beliefs. The distributions of ideologies and party identifications within the analytical sample are plotted in **Figure 11** and **Figure 12**, respectively. Those figures show that the distributions of their respective characteristics do not differ greatly across the two model samples. The primary exception to that rule is the category of people which are "Not Sure" about the identification at hand, ideological or partisan. Those people are largely dropped from the sample through the process of dropping people who were "Not Sure" on the political knowledge questions; this is logical. Moving past people "Not Sure" about their identifications, **Figure 11** shows that the most common ideological identification in the analytical sample is identifying as a moderate; a close second is identifying as conservative. Just over half as many people identify as liberal, followed by roughly 7.5% of the analytical sample identifying "strongly" as conservative and 6.8% as strongly liberal. 2.4% of people included in the analytical sample were not sure about their ideologies. Based on those percentages, it is clear that many more people in the analytical sample identify as conservative than as liberal. However, **Figure 12** shows that, contrary to what a conservative ideological balance might suggest, there are actually more Democrats in the analytical sample than Republicans.

Figure 11: *Ideological Distribution by Model Sample*

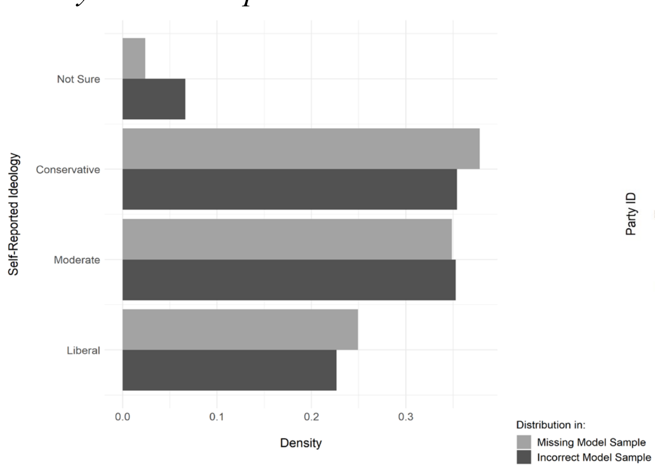
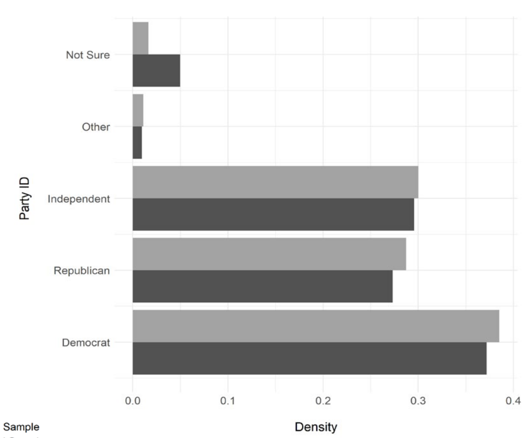
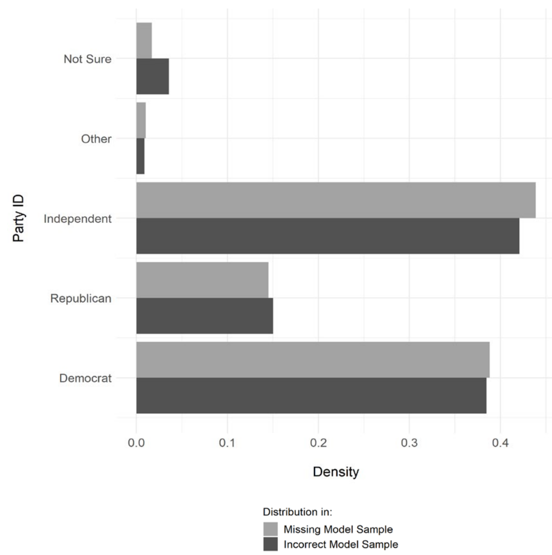


Figure 12: *Distribution of Party Identifications by Model Sample*



The explanation for these two seemingly contradictory trends becomes clear upon inspection of **Figure 13**, which presents the breakdown of party identification among people who described themselves as ideological moderates. As that figure shows, ideological moderates in the sample were over two and a half times as likely to identify as an Independent or a Democrat than to identify as a Republican. This pattern serves as evidence of people's complicated relationships with party ID and ideology, as mentioned in Chapter Three above.

Figure 13: *Party IDs of Self-Reported Ideological Moderates*



2. Regression of Voting on Social Media News Use

Table 9 contains partial output from regressions (1) through (6) from Section 3.6 above, with results given as odds ratios (exponentiated coefficients). (Full regression results are included in Appendix B, **Table 18**). From **Table 9**, two primary conclusions can be drawn: models with state fixed effects strictly dominate those without, and mediated models using the "Incorrect" scores perform better than those using the "Missing" scores. Support for these two conclusions is drawn from *p*-values that result from applying Hosmer-Lemeshow tests to each model. In that test, one of many that can be used to evaluate model goodness of fit, a small *p*-value suggests poor fit; a large value does not necessarily indicate good fit, rather a lack of evidence to *prove* a bad fit – perhaps due to a lack of power/small sample size (Hosmer, Lemeshow, & Sturdivant 2013, 157-169; Archer & Lemeshow 2006).

Table 9: *Exponentiated Coefficients on Treatment and Mediating Variables, Logit Regressions of Voting on Social Media Use*

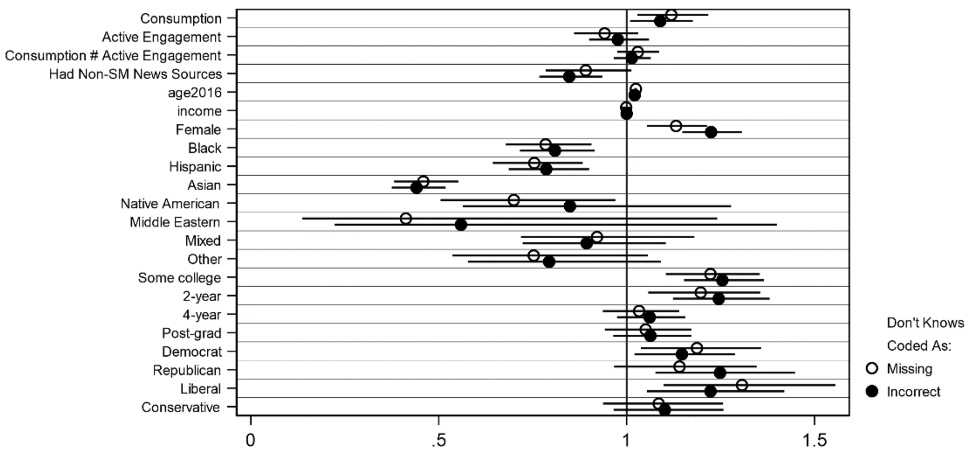
Model Number	**1**	**2**	**3**	**4**	**5**	**6**
State Fixed Effects		✓			✓	✓
Mediation Variables			✓	✓	✓	✓
Treatment of "Not Sure" Responses	N/A	N/A	Missing	Incorrect	Missing	Incorrect
Observations	47,400	47,400	35,120	46,359	35,120	46,359
Hosmer-Lemeshow Test *p*-Value	0.0363	0.2776	0.0021	0.1446	0.0048	0.2276
Consumption	1.135***	1.140***	1.118**	1.086*	1.119**	1.089*
Active Engagement	0.964	0.965	0.936	0.975	0.941	0.976
Consumption * Active Engagement	1.027	1.024	1.034	1.017	1.03	1.014
General Knowledge			1.683***	1.356***	1.660***	1.348***
Personal Knowledge			3.160***	2.684***	3.217***	2.740***
…More Variables…	…	…	…	…	…	…

Exponentiated coefficients, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

From the Hosmer-Lemeshow test p-values in **Table 9**, we can see that models 1, 3, and 5 are poor fits for the data, while we have no evidence to support a similar conclusion for models 2, 4, and 6. These are: the unmediated model with state fixed effects, the "Incorrect" mediated model without stated fixed effects, and the "Incorrect" mediated model with state fixed effects. Therefore, we can safely conclude that the "Incorrect" mediating variables provide the better model. We can also generally assume that including state fixed effects improves the model; for example, within the models which do not include the mediating variables (1 and 2), the model without state fixed effects is shown to be a very poor fit, while the model with state fixed effects is not eliminated. Therefore, models (2) and (6) appear to be the best models for further analysis. While their large *p*-values do not automatically mean that the models are *good* fits, the very large sample sizes reduce the likelihood that the large *p*-values are simply the result of a lack of statistical power to prove model misspecification.

Figure 14 provides another way to compare two of the models from **Table 9**: models 5 and 6. In that figure, we can see that the method for calculating political knowledge affects few variables in the analysis. The variable most significantly affected is female – this is unsurprising, as the main distinction between the two models is whether they drop people who responded "Not Sure" to a significant number of knowledge questions, and those people were two times more likely to be female than male.

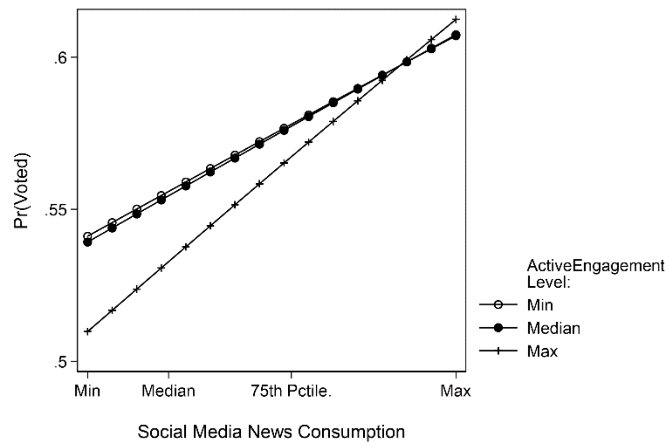
Figure 14: *Odds Ratios Across Mediated Models of Voting (Comparing Methods for Calculating Political Knowledge)*



Both models include state fixed effects; odds ratios for state indicators and for interactions between party ID and ideology not depicted here.

Moving forward with models (2) and (6), the specific effects of variables can be examined. The statistically-significant exponentiated coefficient on Consumption in Model (2) provides evidence that there is in fact a significant, positive effect of consuming news over social media on likelihood of voting, as posited in **H1**. People who consumed news through social media were more likely to vote. Interestingly, however, the statistically-insignificant exponentiated coefficients on ActiveEngagement and on the interaction between Consumption and ActiveEngagement suggest that actively engaging with content has no statistically meaningful impact on likelihood of voting. Nonetheless, there is a pattern, as seen in **Figure 15**:

Figure 15: *Predictive Margins of Consumption on Voting, by Level of Active Engagement*



more ActiveEngagement seems to reduce likelihood of voting (seen in the 0.964 odds ratio on ActiveEngagement), though this effect varies by level of Consumption. As Consumption level increases, ActiveEngagement's effect decreases in size. People who scored less than the maximum value on Consumption are generally less likely to vote; they are even *less* likely to vote if they are low consumers *but* highly-active sharers. These differences cannot be stated with absolute certainty, due to the statistical insignificance of the interaction term; technically, the trends are statistically indistinguishable from one another. However, graphic representation of the trends suggests a general pattern.

Most other variables in model (2) are also statistically significant, as seen in Appendix B, **Table 18**: age, income, educational attainment, race, party ID, ideology, interactions between part ID and ideology, and state. The only confounds not statistically significant in model (2) are OtherNewsUse and sex. The exponentiated coefficients on the statistically-significant confounds in the model suggest that (separately from each other) being older, having more income, having more than a high-school education, being white, identifying with a party, and identifying with a specific ideology each predicts increased likelihood of voting. Additionally, living in Colorado, Indiana, Iowa, Maine, Michigan, Minnesota, Missouri, Nevada, North Carolina, Ohio, Oregon, Pennsylvania, or Washington is predicted to make one more likely to vote in 2016 *than someone living in California*. Finally, having a party ID and ideology that do not traditionally align (i.e.

Democrat # Conservative and Republican # Liberal) predicts lower likelihood of voting than identifying as an Independent # Moderate, while having a party ID but being "Not Sure" about one's ideology predicts dramatically higher likelihood of voting than identifying as an independent moderate.

Adding the variables for political knowledge to the model leads Consumption's effect to all but disappear. In that model (6), both political knowledge scores are strongly predictive of voting (PersonalKnowledge more so than GeneralKnowledge), while Consumption's effect is reduced in both magnitude and statistical significance. ActiveEngagement's effect and that of the interaction term remain statistically insignificant, though both also revert closer to 1 (i.e. reduce in magnitude). Taken together, this provides strong evidence for mediation of social media news use's effect on voting by political knowledge, as posited in **H2**. However, this mediation is only partial, as Consumption retains some statistical significance – if that term had become wholly insignificant, I could then claim complete mediation of news-related social media use's effect on voting.

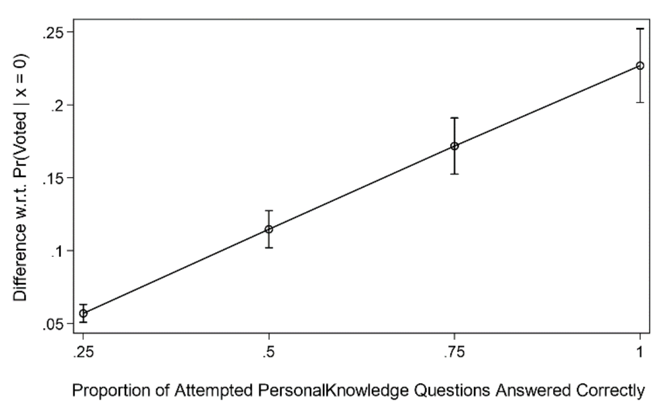
In model 6, the pattern in voting as relates to Consumption and ActiveEngagement is still similar to that found in model 2; consumption predicts more voting, and engagement predicts less voting, though that negative effect is reduced as consumption increases and neither ActiveEngagement's nor the interaction term's effect is statistically significant. Now, however, the magnitude of Consumption's effect has decreased, and an increase in Consumption score has slightly less of an impact on likelihood of voting than when the knowledge scores were not present in the regression.

Holding everything else in the equation constant (active engagement level, political knowledge, other news use, age, gender, education, race, and strength of beliefs) someone with the minimum possible Consumption score (did not consume any news over social media) was predicted to have a 58.95% likelihood of voting. Someone at the median value of Consumption was predicted to have a 60.2% likelihood of voting, and someone at the max value of Consumption was predicted to have a 64.6% likelihood of voting – so consuming news over social media does increase one's likelihood of voting. Both the median and the max predictions are statistically different from the prediction at the minimum, though only to the $p < 0.05$ level.

There is also a significant difference in voting between people with varying levels of political knowledge. This follows, if much of Consumption's effect is mediated through knowledge, as indicated. People who correctly answered neither of the GeneralKnowledge questions were predicted to vote 52.05% of the time; compared to 58.5% of the time for people who were right on each question they attempted.

An even larger jump is found when calculating predictive margins for PersonalKnowledge: people who attempted at least one question but correctly answered none of them were predicted to vote 38.5% of the time, compared to 51.2% of the time for people who got half of their attempted questions right and 63.6% of the time for people who aced all of the questions they attempted. In other words, people who got every question right that they attempted were 65% more likely to vote than people who attempted at least one question but got none right (a difference of almost 20 percentage points, as seen in **Figure 16**).

Figure 16: *Contrasts of Predictive Margins of PersonalKnowledge on Voting*



These differences are so stark, that they must be driven by differences in more than just knowledge scores themselves, because the scores' odds ratios, though large, are nowhere near large enough to account for overall twelve and twenty-five percentage-point differences. The next set of regressions (7 and 8) will investigate the differences in people by knowledge level.

However, before moving on to those regressions, I quickly look at the remaining variables in regression (6). Specifically, I look at the effects of state and partisanship. First, I find that – as stated before – the state in which one lived was a statistically significant predictor of likelihood of voting. This comports with reality; state-level turnout varies from the low 40s to upper 70s. Interestingly, the effect of state in this regression is inversely related to actual turnout, as depicted in **Figure 17**; where turnout was already highest, social media news consumption made the least difference in it. Direct comparison of those trends is plotted in **Figure 18** below, in which we see that turnout and the treatment effect are, in fact, inversely related (following, generally, a pattern of exponential decay).

Figure 17: *Heatmaps of Statewide Turnout and State-Level Contrasts in Predicted Turnout by Level of Social Media News Consumption*

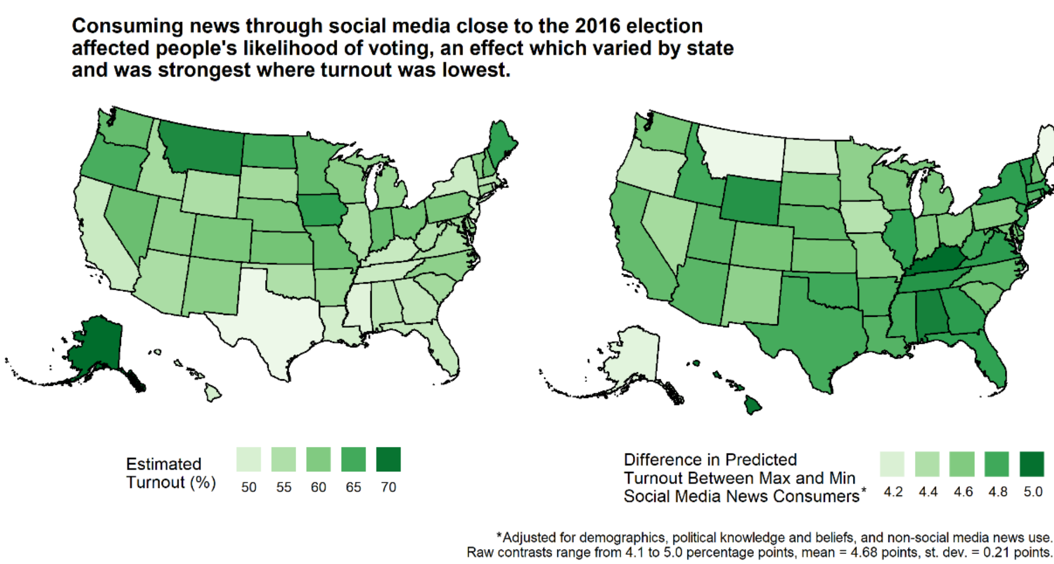
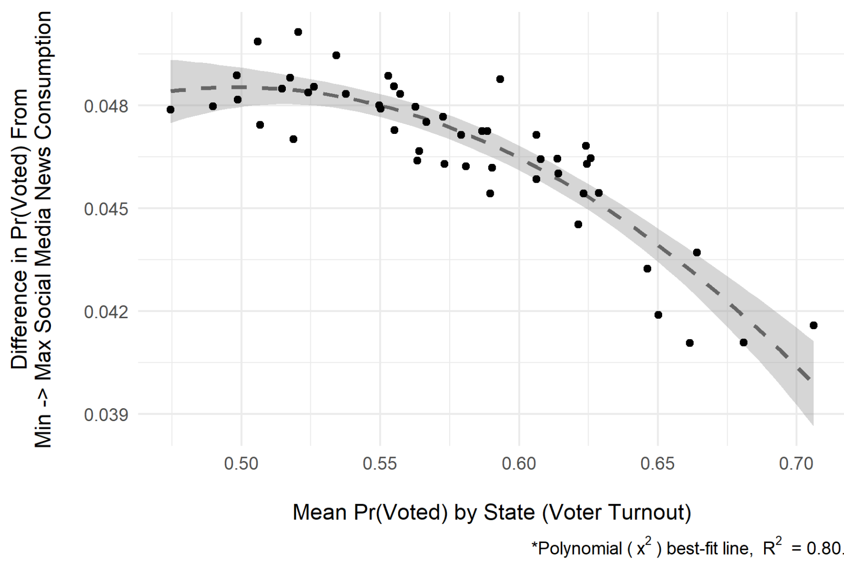


Figure 18: *Scatterplot of Statewide Turnout against State-Level Contrasts in Predicted Turnout by Level of Social Media News Consumption*



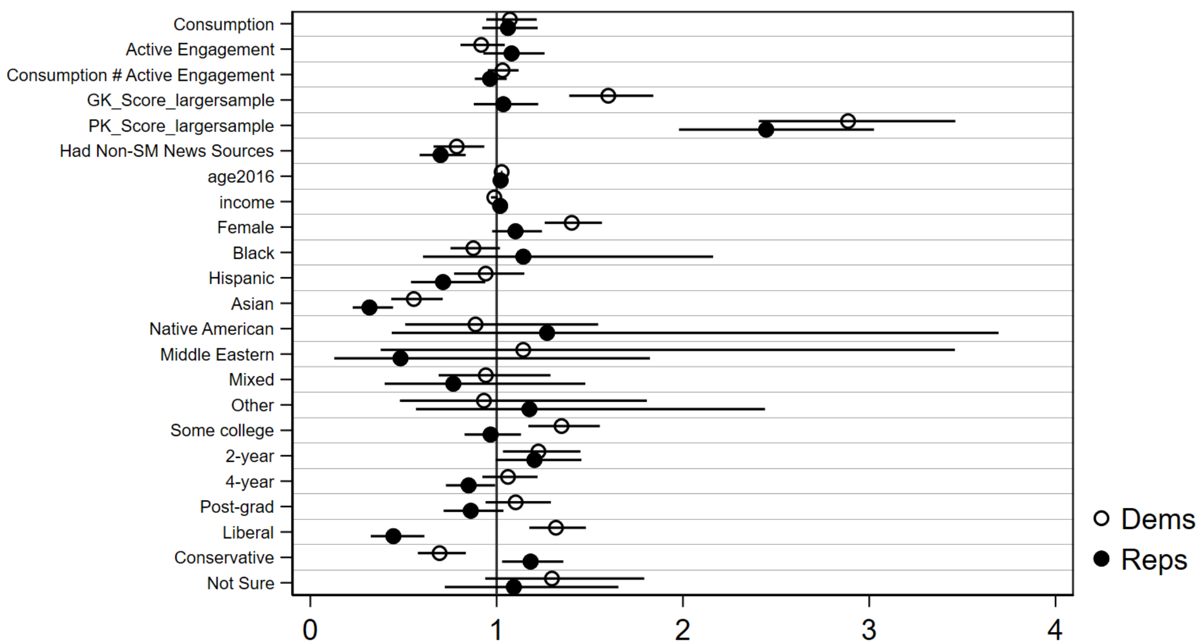
Most states were generally like one another as regards turnout; 34 states fell within ± 1 standard deviation of the mean state-level prediction turnout (57.6% ± 5.33%). Among those 34 states, the mean treatment effect was +4.71 points. Among the six states with turnout more than one standard deviation below the mean, the mean treatment effect was 4.21 points – and among the ten states with turnout more than one standard deviation above the mean, the mean treatment effect was 4.85 points. So, the treatment effect did vary by state, if only slightly.

In addition to investigating state-level effects, I also looked at split models for Democrats and Republicans, separately. To do so, I ran model (6) for subpopulations of the survey sample, based on party ID – resulting in two sub-regressions, in which all respondents in each regression were either Democrats or Republicans, specifically. Doing so allows examination of whether patterns and relationships between variables are functioning differently for partisans of different colors.

The results of those regressions are plotted in **Figure 19**, below, which shows that several variables perform quite differently between the two subpopulations. Perhaps most importantly, the knowledge variables' effects were starkly different within the Democratic and within the

Republican populations. Both knowledge scores had a stronger effect on voting among Democrats than among Republicans, suggesting that more-knowledgeable Democrats were more likely to vote than less-knowledgeable Democrats, but that Republicans were more similarly-likely to vote regardless of knowledge level. This is particularly true of GeneralKnowledge, in which the effect in the Republican model is statistically insignificant. For PersonalKnowledge, more-knowledgeable Republicans are more likely to vote than less-knowledgeable Republicans, though that gap is smaller than it is for Democrats.

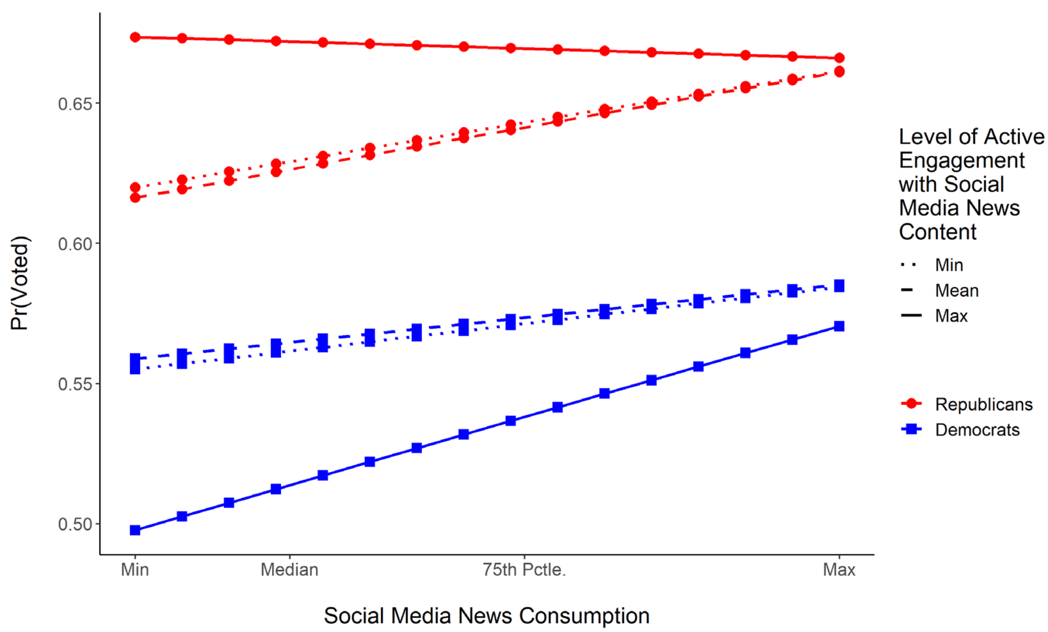
Figure 19*: Odds Ratios Across Split Models of Voting by Party ID (Comparing Models Limited to only Democrats and only Republicans)*



Additionally, being female had a stronger effect on voting among Democrats than Republicans, though this may be influenced by the underlying gender characteristics of those two populations. Racial variables also may be experiencing similar patterns. Ideological "alignment" clearly predicted higher voting, while "misalignment" predicted lower voting; liberal Democrats voted more than liberal Independents, who voted more than liberal Republicans, and conservative Republicans voted more than conservative Independents, who voted more than conservative Democrats.

Finally, we can see that the social media variables perform differently in these two sub-regressions. For people of both parties, consuming social media news predicted higher voting. However, Active Engagement's moderation effect on Consumption works completely differently. Democrats who were active social media sharers were predicted to vote less than their less-active peers, while actively-engaged Republicans were more likely to vote than their less-active peers.

Figure 20: *Predictive Margins of Social Media News Consumption on Voting by Level of Active Engagement, across Separate Models for Democrats and Republicans*



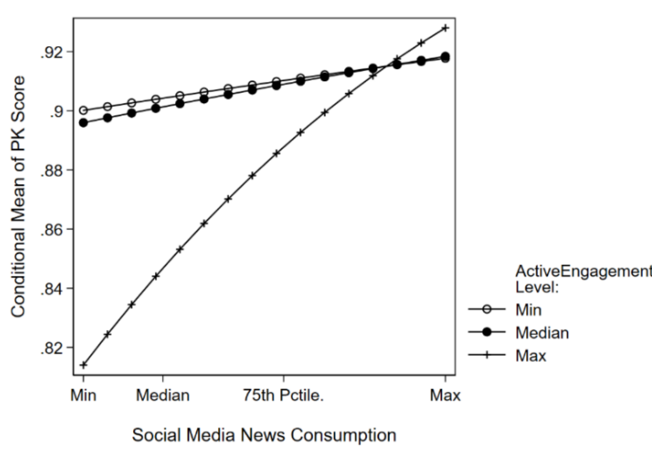
Nationwide, Democrats' and Republicans' likelihoods of voting in 2016 were differently impacted by their news-related social media use around the time of the election.

3. Regressions of Political Knowledge on Social Media News Use

Finally, **Table 19** in Appendix B presents the results of regressions (7) and (8). As discussed in the section immediately above, there seem to be additional, significant differences among people with different political knowledge scores other than just their political knowledge score. The results presented in this section give insight into what those differences may be. As **Table 19** shows, all but one variable used in both regressions are statistically significant predictors of knowledge levels, the exception being OtherNewsUse. Consuming news through social media was found to be a positive predictor of knowledge, as expected based on the mediation effect found in regressions (2) and (6) above. Additionally, as was the case when the outcome was voting, active engagement with social media news was a negative predictor of knowledge, when controlling for consumption. The interaction term was also again a positive predictor, indicating that ActiveEngagement's negative effect reduces in magnitude as Consumption increases.

The relationship between Consumption and ActiveEngagement is seen in **Figure 21**: for more than half of people, being more-highly engaged with social media news content actually predicts *lower* personal political knowledge. It is only at the highest levels of Consumption that one's level of active engagement makes no difference on personal political knowledge. These findings provide evidence against **H3**, and contradict the theory put forward at

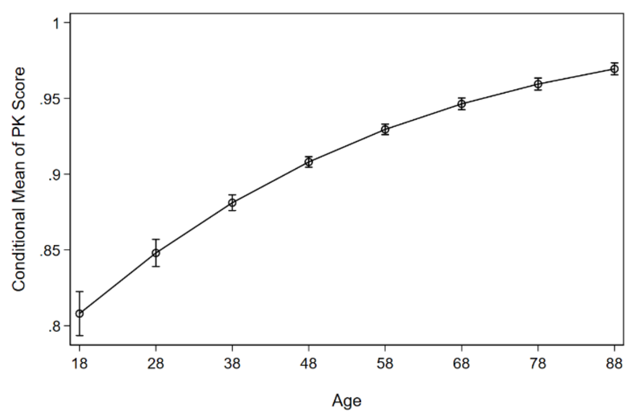
Figure 21: *Predictive Margins of Social Media News Consumption on Personal Political Knowledge, by Level of Active Engagement*



the conclusion of Jung, Kim, and Gil de Zúñiga (2011) – further discussion can be found in the Conclusions chapter below.

Among the demographic predictors, being older, higher income, and more educated are all positive predictors of political knowledge. For example, **Figure 22** depicts respondents' predicted accuracy on PersonalKnowledge questions by age; octogenarians are generally 20% more accurate on PersonalKnowledge questions than people just entering young adulthood. Additionally, being White predicts higher political knowledge than being Black, Hispanic, or Asian – though being Middle Eastern or self-identifying as "Other" predicts higher political knowledge than identifying as White.

Figure 22: *Predictive Margins of Age on PersonalKnowledge Score*



Notably, the results presented in **Table 19** indicate a lingering gender difference in political knowledge; being male predicts a higher score on both general and personal political knowledge scores. However, that difference is less for the personal political knowledge score than for the general political knowledge score (an allowable comparison because both scores have the same range [0,1]). Women are generally 5.8% less accurate on GeneralKnowledge than men, but only 2.3% less accurate than men on PersonalKnowledge. This once again provides evidence that localizing/personalizing political knowledge questions is one way to reduce gender bias in political knowledge scores.

Finally, both party ID and ideology are predictors of political knowledge. Taken individually, knowledge seems not to differ greatly among different ideologies; on party ID, being an Independent does appear to predict higher knowledge than aligning with a party or being not sure.

When the two trends are taken together, as in **Figure 23**, a pattern like that found in voting emerges. People with traditionally-aligned party IDs and ideologies (e.g. conservative Republicans) are more knowledgeable than people with traditionally-unaligned IDs and ideologies (e.g. liberal Republicans). Ideologically-aligned partisans on the whole also tend to know more than Independent moderates, a difference generally significant at the $p < 0.05$ level.

Figure 23: *Matrix of Predictive Margins of Personal Political Knowledge by Party ID and Ideology*

		Ideology					
		VL	Lib.	Mod.	Cons.	VC	NS
PID	SD	92.8	92.8	87.8	80.9	78.4	83.9
	Dem.	96.3	92.1	90.3	81.1	76.6	87.5
	Ind.	91.2	86.9	91.1	88.7	92.7	78.3
	Rep.	77.3	81.1	89.6	91.9	96.2	74.4
	SR	78.3	74.1	84.9	92.8	92.5	86.3
	NS	89.3	72.3	91.7	81.4	60.9	82.9

Conclusions

The 2016 election was novel in many ways, including the use of social media by voters, media, candidates, and foreign influences. Based on the findings of this thesis, it appears that people's interaction with political news through social media around the time of the election did partially predict whether they would participate in it. In this chapter, I review the results from Chapter Four above and synthesize key findings. In the first section below, I specifically evaluate each hypothesis that was formulated at the beginning of this thesis. Then, I consider the implications of this thesis' findings both for the body of literature within which this thesis exists and for the stakeholders who are invested in the relationship between social media and voting. Finally, I qualify the findings of this thesis by enumerating its limitations and, mostly based on those limitations, I recommend several directions for future research.

1. Evaluation of Hypotheses

In analysis, evidence was found to support **H1**; after controlling for demographic factors, political beliefs, and getting news from other non-social media sources, consuming political news content over social media in the period shortly before the 2016 election was found to increase the likelihood that someone voted. If someone consumed social media news content, whether they then shared it, commented on it, or made their own post made no difference on their likelihood of voting. However, if someone did not consume content, *but* nonetheless otherwise actively engaged with content – i.e. shared it, commented on it, or created their own – they were found to be less likely to vote than people who neither consumed nor engaged with content.

Evidence was also found to support **H2**; the effect of social media news use was found to be partially mediated by political knowledge. Primary evidence of this is provided by the exponentiated coefficients on Consumption, ActiveEngagement, and their interaction across regressions (2) and (6). Upon the addition of political knowledge to the regression, the effects of social media news use are reduced in both magnitude and significance. Secondary support for a mediatory pathway is provided by regressions (7) and (8); social media news use was found to be a significant predictor of political knowledge level, which is necessary for political knowledge to be a mediator of social media news use's effect on voting.

Evidence was also found to support **H3**; most significantly, I found that Democrats' odds of voting were more impacted by how much they know about politics, while Republicans' likelihoods of voting were less conditional on how much they knew about politics. This affects the social media → knowledge pathway as well; Democrats who were active social media sharers were predicted to vote less than their less-active peers, while actively-engaged Republicans were more likely to vote than their less-active peers.

Finally, both steps taken in attempt to reduce gender bias in measured political knowledge were found to help achieve that goal, as posited by **H4** and **H5**. Limiting analysis to only *attempted answers* reduced the observed effect of gender on political knowledge, and the knowledge score which examined people's awareness of their own representation was also less affected by gender than the score which measured people's knowledge of larger political balances of power. However, it should be noted that the first step – limiting analysis – removed over 10,000 respondents from analysis and should therefore be taken with a grain of salt.

2. Limitations of Methods

Several aspects of the research design, methods, and data used in this thesis should be noted for their impacts on its findings and their generalizability. First, as discussed in Chapter Three, the digital sampling frame used to collect the survey data limits the external validity of the findings in this thesis. Because only relatively-digitally-savvy people could participate in the survey, the sample is almost inevitably not representative of people who do not frequently use computers and the internet. As has been seen, this segment of the population is growing ever smaller; still, it cannot be discounted.

Secondly, the secondary nature of the data limited the questions that could be used in analysis. The main constraint imposed as a result of this was on the questions used to evaluate respondents' social media behavior. In the CCES, survey-takers were asked simply whether they had engaged in certain activities in the last twenty-four hours. As discussed above, this may actually be a more accurate measure than asking people to estimate the frequency of their use over any given time period. However, the standard in the field of social media study is to use measures of frequency, rather than a binary yes/no for having engaged at all. While this distinction does not limit the findings of this thesis per se, frequency measures likely would have had greater internal variation than the heavily bottom-weighted social media measures used in analysis.

The nature of CCES data also limit the findings of this thesis in that they are observational, as are most survey data. Findings supporting the hypothesis that people who use social media for news are more knowledgeable and more likely to vote would be much stronger if gathered through an experiment, in which I was able to manipulate people's social media use. Without this intervention, the causal claims that I have made are more tenuous. Some support for the causal ordering that I have posited here is provided by previous work which tested each possible causal ordering and found the social media news use → knowledge → voting order to be the best fit for the data (Jung, Kim, & Gil de Zúñiga 2011). However, those findings are most relevant to only the election they studied (2008), and the most robust study of news-related social media use in the 2016 election would include some aspect of intervention/manipulation itself.

3. Implications of Findings

Many of the findings of this thesis are in concordance with prior research in the field, which suggests that social media's effect on participation in the 2016 election did not differ tremendously from its effect on participation in previous contexts. Primarily, the findings that social media news consumption predicts higher likelihood of voting and that that effect was found to be mediated by political knowledge build upon the foundation established by Kenski and Stroud (2006), Shah et al. (2005), and Jung, Kim, and Gil de Zúñiga (2011).

Other evidence found in this thesis also adds to previous research regarding gender bias in analytical methods for evaluating political knowledge; the process of excluding "don't know" answers from analysis and using questions that are more personalized to the individual respondent was shown to produce a less gender-biased score for political knowledge – a combination of steps like those taken by Fortin-Rittberger (2016) and Miller (2018).

The primary finding of this thesis, that using social media for political news impacts voting, bears directly on not only policymakers, but also media companies and social media platforms themselves, all of whom are, at this moment, attempting to sort out for themselves the extent to which social media impacts elections. The indication, in this thesis, that consuming news content through social media is associated with higher political knowledge is a positive note, signaling that people *are* paying attention to the news stories that come across their feeds. However, that same fact should also be taken as a sign of increased urgency for the "fake news" debate. While this thesis was able to test for the normatively-positive effect on knowledge of people consuming social media news content, it was not able to test for adverse effects, because it had no way to distinguish between people's exposure to "real" and "fake" news. Therefore, while people did benefit in at least one identifiable way from using social media for political news, there may be myriad other ways that they were impacted that could not be not tested for in this thesis, such as an increased likelihood to believe false stories about the election.

Additionally, the second main finding of this thesis, that people who do not consume news content through social media – yet still *share* such content – are both less politically knowledgeable and less civically engaged (through voting) is a cause for concern. It should be noted that this finding goes against the theory put forward at the end of Jung, Kim, and Gil de Zúñiga (2011). That paper found that, although social media news consumption does predict both political knowledge and voting, its effect on political knowledge is practically small. The authors theorize in their conclusions that, in order to "acquire" political knowledge, more intense processing of information may be required than just consuming it (423). That theory would suggest that, in this thesis, ActiveEngagement would be a positive predictor of political knowledge, and indeed a stronger one than Consumption.

However, as has been seen, such is not the case. When people both consume and actively engage with content, they gained no more knowledge than if they had only consumed content, and when people *only* actively engaged and did not consume, they actually were predicted to have less knowledge than if they had done neither. Barring a measurement or modeling error, the logical conclusion of this finding is that a non-trivial portion of the people sharing political news content on social media are less knowledgeable and less involved in the formal political process than average. When taken in combination with several sets of investigators' findings regarding social media "opinion leaders'" capacity to influence vast numbers of other users with their posts (Morrell 2012; Gil de Zúñiga et al. 2014; Allcott & Gentzkow 2017; Weeks, Ardèvol-Abreu, & Gil de Zúñiga 2017), that possibility becomes a strong argument for further refinement of social media platforms' processes regarding news, such as the step already taken to identify "trusted" or locally-relevant news sources on Facebook (Mosseri 2018).

4. Suggestions for Future Research

The questions asked and methods used in conducting research into social media and its effects are constantly in flux, due to the ever-changing nature of the platform. Since the 2016 election, that has especially been the case, since social media sites have started to be clued into the extent of their influence. This thesis contributes to the literature that illuminates social media's influence – but there are extensions of this thesis' analysis that could take its findings even further.

First and foremost, future research would benefit greatly from use of an experimental, rather than an observational, design. The claims of causality made in this thesis would be much stronger if they were the result of intervention. However, this approach can also be difficult in the context of a specific election. For example, I, in 2019, could not perform experimental research about the 2016 election – nor might any researchers at the time of the 2016 election have anticipated the full extent of the role that social media would play in that election and therefore perform such an experiment. Going forward, however, that option always remains.

Secondly, as has been mentioned several times throughout, future research would benefit greatly from the ability to distinguish between “real” news and “fake” news stories when examining the effect of social media use on political knowledge and voting. Studies have begun to develop methods for distinguishing between the two types; Silverman (2016) used post-level data from Facebook to draw conclusions at the content level (though subsequent authors have called those methods into question – Rinehart 2017). Perhaps an extension or modification of his method could be used to examine Facebook data at the user level – tying together the individual-level analysis of this thesis with “fake news” exposure information from Facebook.

Additionally, future research would benefit from the ability to include personality among its control variables. In this thesis, personality was left to reside in the error terms of regressions; however, as reviewed in Chapter Two, it is a very useful predictor of social media behavior, and it may also be tied to voting, though the connection has been made less rigorously (Samek 2016). The connection between various personality traits and political knowledge gain from social media news use would be of great interest, as well. For example, would openness to experience (one of the Big Five personality traits) affect the amount of political knowledge gleaned from consuming news content? It seems reasonable to expect so.

Future research into political knowledge should also, as the trend has been, continue to refine the questions used to evaluate that construct. The more-personalized questions used in this thesis did prove to be less subject to gender bias than the broader, nationally-scaled questions; however, other questions have proved to be even more effective in reducing gender bias, such as the gender-relevant questions used by Dolan (2011). Inclusion of such knowledge questions in large, national data sets like the CCES would certainly be a boon to that field of research.

Finally, future work might consider a direct extension of this thesis with an added a time dimension: apply the methods of this thesis to not only 2016 CCES data, but also CCES data from the 2008 and 2012 elections, and potentially the midterm elections in-between. Such longitudinal analysis would allow tracking of news-related social media use's effect over the time period in which social media were developed and grew to prominence.

References

Agadjanian, A. (2018a, February 10). Turnout Underestimates and Voter File Match Rate Problems in the 2016 CCES [Blog post]. https://agadjanianpolitics.wordpress.com/2018/02/10/turnout-underestimates-and-voter-file-match-rate-problems-in-the-2016-cces/

Agadjanian, A. (2018b, May 6). Over-Time State and National Bias in CCES Vote Validated Turnout Rates [Blog post]. https://agadjanianpolitics.wordpress.com/2018/05/06/over-time-state-and-national-bias-in-cces-vote-validated-turnout-rates/

Allcott, H. & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives 31*(2), 211-236. Retrieved from https://www.doi.org/10.1257/jep.31.2.211

Ansolabehere, S. & Hersh, E. (2012, August 27). Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate. *Political Analysis* 20, 437-459. Retrieved from https://doi.org/10.1093/pan/mps023

Ansolabehere, S. & Schaffner, B. (2017). Cooperative Congressional Election Survey 2016: Common Content. [Computer File] Release 2: August 4, 2017. Cambridge, MA: Harvard University [producer] http://cces.gov.harvard.edu

Ansolabehere, S., Schaffner, B., & Luks, S. (2017, August). Guide to the 2016 Cooperative Congressional Election Survey. Retrieved from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910/DVN/GDF6Z0

Annenberg Public Policy Center (APPC). (2014). Americans know surprisingly little about their government, survey finds. Retrieved from https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2018/03/Civics-survey-press-release-09-17-2014-for-PR-Newswire.pdf

Archer, K.J. & Lemeshow, S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The State Journal 6*(1), 97-105. Retrieved from https://www.stata-journal.com/sjpdf.html?articlenum=st0099

Bagdikian, B.H. (1983). *The Media Monopoly.* Boston, MA: Beacon Press.

Bartholomew, D., Knott, M., & Moustaki, I. (2011, June 16). *Latent Variable Models and Factor Analysis: A Unified Approach, 3rd Edition.* Hoboken, NJ: John Wiley & Sons, Inc.

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford: Oxford University Press.

Bernstein, R., Chadha, A., & Montjoy, R. (2001). Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly 65*(1), 22-44. Retrieved from https://doi.org/10.1086/320036

Brennan, J. (2016). *Against Democracy*. Princeton, NJ: Princeton University Press.

Buckingham, D. (2015). Do We Really Need Media Education 2.0? Teaching Media in the Age of Participatory Culture. In T.-B. Lin et al. (Eds.) *New Media and Learning in the 21st Century*. Singapore: Springer Science + Business Media. http://dx.doi.org/10.1007/978-981-287-326-2_2

Correa, T., Hinsley, A.W., & Gil de Zúñiga, H. (2010. Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior 26*, 247-253. Retrieved from: https://doi.org/10.1016/j.chb.2009.09.003

Cuevas-Molina, I. (2017). Voter Turnout Overreports: Measurement, Modeling and Deception. *Doctoral Dissertations* 958. Retrieved from https://scholarworks.umass.edu/dissertations_2/958

Delli Carpini, M.X. & Keeter, S. (1996). What Americans Know about Politics and Why It Matters. New Haven, CT: Yale University Press. Retrieved from https://www.jstor.org/stable/j.ctt1cc2kv1

DiStefano, C., Zhu, M., & Mîndrilă, D. (2009, October). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research & Evaluation 14*(20). Retrieved from https://pareonline.net/pdf/v14n20.pdf

Dolan, K. (2011, January 14). Do Women and Men Know Different Things? Measuring Gender Differences in Political Knowledge. *The Journal of Politics 73*(1), 97-107. Retrieved from https://www.jstor.org/stable/10.1017/s0022381610000897

Duggan, M. & Brenner, J. (2013). The Demographics of Social Media Users – 2012. Retrieved from http://boletines.prisadigital.com/PIP_SocialMediaUsers.pdf

Fortin-Rittberger, J. (2016). Cross-National Gender Gaps in Political Knowledge: How Much Is Due to Context? *Political Research Quarterly 69*(3), 391-402. Retrieved from https://doi.org/10.1177/1065912916642867

Freiling, N. (2017, May 31). Just 37% of Americans can name their Representative. Retrieved from https://www.haveninsights.com/just-37-percent-name-representative/

Garand, J.C., Guynan, E., & Fournet, M. (2004). The Gender Gap and Political Knowledge: Men and Women in National and State Politics. Paper presented at the 2004 annual meeting of the Southern Political Science Association, New Orleans, LA, January 8-10, 2005. Retrieved from http://citation.allacademic.com//meta/p_mla_apa_research_citation/0/6/7/9/9/pages67995/p67995-1.php

Gil de Zúñiga, H., Jung., N., & Valenzuela, S. (2012). Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer Mediated Communication 17*, 319-336. Retrieved from https://doi.org/10.1111/j.1083-6101.2012.01574.x

Gil de Zúñiga, H., Molyneux, L., & Zheng, P. (2014). Social Media, Political Expression, and Political Participation: Panel Analysis of Lagged and Concurrent Relationships. *Journal of Communication 64*, 612-634. Retrieved from https://doi.org/10.111/jcom.12103

Goldberg, L.R., Sweeney, D., Merenda, P.F., & Hughes Jr., J.E. (1998). Demographic Variables and Personality: The Effects of Gender, Age, Education, and Ethnic/Racial Status on Self-Descriptions of Personality Attributes. *Personality and Individual Differences 24*(3), 393-403. Retrieved from https://projects.ori.org/lrg/PDFs_papers/Demographic%20Variables%20and%20Personality.pdf

Grimmer, J., Hersh, E., Meredith, M., Mummolo, J., & Nall, C. (2018, April 18). Obstacles to Estimating Voter ID Laws' Effect on Turnout [Scholarly Dialogue]. *The Journal of Politics 80*(3), 1045-1051. Retrieved from http://dx.doi.org/10.1086/696618

Groshek, J. & Koc-Michalska, K. (2017). Helping populism win? Social media use, filter bubbles, and support for populist presidential candidates in the 2016 US election campaign. *Information, Communication & Society* 20(9), 1389-1407. Retrieved from https://doi.org/10.1080/1369118X.2017.1329334

Habing, B. (2003, October 15). Exploratory Factor Analysis [Handout]. Retrieved from http://people.stat.sc.edu/habing/courses/530EFA.pdf

Hair Jr., J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate Data Analysis (5th Ed.)*. Upper Saddle River, NJ: Prentice Hall.

Hawkins, C.B. & Nosek, B.A. (2012, November). Motivated independence? Implicit party identity predicts political judgments among self-proclaimed Independents. *Personality & Social Psychology Bulletin 38*(11), 1437-1452. Retrieved from https://doi.org/10.1177/0146167212452313

Hitlin, P. (2018, September). Internet, social media use and device ownership in U.S. have plateaued after years of growth. Retrieved from http://www.pewresearch.org/fact-tank/2018/09/28/internet-social-media-use-and-device-ownership-in-u-s-have-plateaued-after-years-of-growth/

Hosmer, Jr., D.W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied Logistic Regression*. Hoboken, NJ: John Wiley & Sons. https://bit.ly/2XNqWWe

Jamieson, K.H. (2018). *Cyberwar: How Russian Hackers and Trolls Helped Elect a President; What We Don't, Can't, and Do Know.* Oxford: Oxford University Press.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: Understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analyses.* New York, NY: ACM.

Jung, N., Kim, Y., & Gil de Zúñiga, H. (2011). The Mediating Role of Knowledge and Efficacy in the Effects of Communication on Political Participation. *Mass Communication and Society 14*(4), 407-430. Retrieved from https://doi.dorg/10.1080/15205436.2010.496135

Kaplan, R.L. (2002). *Politics and the American Press: The Rise of Objectivity, 1865-1920.* Cambridge, UK: Cambridge University Press. Retrieved from https://ebookcentral-proquest-com.libproxy.lib.unc.edu/lib/unc/detail.action?docID=3004523

Kenski, K. (2006). The Gender Gap in Political Knowledge: Do Women Know Less Than Men About Politics? [Dissertation]. University of Pennsylvania, *ProQuest Dissertations Publishing*. Retrieved from https://search-proquest-com.libproxy.lib.unc.edu/docview/305258911

Kenski, K. & Stroud, N.J. (2006, June). Connections Between Internet Use and Political Efficacy, Knowledge, and Participation. *Journal of Broadcasting & Electronic Media 50*(2), 173-192. Retrieved from https://doi.org/10.1207/s15506878jobem5002_1

Khang, H., Han, E-K., Ki, E-J. (2014). Exploring influential social cognitive determinants of social media use. *Computers in Human Behavior 36,* 48-55. http://dx.doi.org/10.1016/j.chb.2014.03.038

Kreiss, D. (2019). From Epistemic to Identity Crisis: Perspectives on the 2016 U.S. Presidential Election. *The International Journal of Press/Politics,* 1-6. Retrieved from https://doi.org/10.1177/1940161219843256

Lang, K. & Lang, G.E. (2002). *Television and Politics*. Piscataway, NJ: Transaction Publishers.

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010, February). Social Media & Mobile Internet Use among Teens and Young Adults. Millenials. [Report]. The Pew Internet & American Life Project. Retrieved from https://eric.ed.gov/?id=ED525056

Leskovec, J., McGlohan, M. Faloutsos, C., Glance, N., & Hurst, M. (2007). Patterns of cascading behavior in large blog graphs. *Proceedings of the SAIM International Conference on Data Mining,* 551-556. Philadelphia, PA: SIAM.

Lin, N. (2008). A network theory of social capital. In D. Castiglione, J.W. van Deth, & G. Wolleb (Eds.), *The Handbook of Social Capital* (50-69). London: Oxford University Press.

Lizotte, M-K. & Sidman, A.H. (20019, June 17). Explaining the Gender Gap in Political Knowledge. *Politics & Gender 5*(2), 127-151. Retrieved from https://doi.org/10.1017/S1743923X09000130

Mandiberg, M. (2012). The Social Media Reader. Retrieved from https://ebookcentral-proquest-com.libproxy.lib.unc.edu

McDonald, M. (n.d.). Voter Turnout Data. *United States Election Project.* Retrieved from http://www.electproject.org/home/voter-turnout/voter-turnout-data

Mosseri, A. (2018, January 19). Helping Ensure News on Facebook is From Trusted Sources [Blog Post]. Retrieved from https://newsroom.fb.com/news/2018/01/trusted-sources/

Miller, M.K. (2018, October 3). Who Knows More About Politics? A Dual Explanation for the Gender Gap. *American Politics Research 47*(1), 174-188. Retrieved from https://doi-org.libproxy.lib.unc.edu/10.1177/1532673X18803890

MIT Election Data & Science Lab. (n.d.) Voter turnout. https://electionlab.mit.edu/research/voter-turnout

Montgomery, K. & Gottlieb-Robles, B. (2006). Youth as e-Citizens: The Internet's Contribution to Civic Engagement. In D. Buckingham & R. Willett (Eds.), *Digital generations: children, young people, and new media* (131-148). Mahwah, NJ: Lawrence Erlbaum Associates.

Moore, K. & McElroy, J.C. (2011). The influence of personality on Facebook usage, wall postings, and regret. *Computers in Human Behavior 28*, 267-274. http://dx.doi.org/10.1016/j.chb.2011.09.009

Morrell, E. (2012). 21ST-CENTURY LITERACIES, CRITICAL MEDIA PEDAGOGIES, AND LANGUAGES ARTS. *The Reading Teacher 66*(4), 300-302. https://www.jstor.org/stable/23321310

Norrander, B. & Wilcox, C. (2008, April 15). The Gender Gap in Ideology. *Political Behavior 30*, 503-523. Retrieved from https://doi.org/10.1007/s11109-008-9061-1

O'Reilly, T. (2007). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. In Mandiberg, M. (Ed.) *The Social Media Reader* (32-52). New York, NY: New York University Press.

Pearl, J. (1998, November 1). Graphs, Causality, and Structural Equation Models. *Sociological Methods & Research 27*(2), 226-284. Retrieved from https://doi.org/10.1177/0049124198027002004

Pew Research Center. (2007, September). Political Knowledge Update. Retrieved from http://www.pewresearch.org/wp-content/uploads/sites/4/legacy-pdf/486.pdf

Pew Research Center. (2018a, February). Internet/Broadband Fact Sheet. http://www.pewinternet.org/fact-sheet/internet-broadband/

Pew Research Center. (2018b, April). The Public, the Political System and American Democracy. Retrieved from http://www.people-press.org/wp-content/uploads/sites/4/2018/04/4-26-2018-Democracy-release1.pdf

Rahn, M. (n.d.). Factor Analysis: A Short Introduction, Part 2-Rotations. Retrieved from https://www.theanalysisfactor.com/rotations-factor-analysis/

Rinehart, W. (2017, January 31). Was There Really A Fake News Epidemic? Three Reasons Why You Shouldn't Rely On The Buzzfeed Analysis Of Fake News [Blog post]. Retrieved from https://medium.com/@willrinehart/was-there-really-a-fake-news-epidemic-e22452d8312e

Rogers, L.T. (2016, June 23). America's Age Told through Population Pyramids. United States Census Bureau, Population Estimates Branch. Retrieved from https://www.census.gov/newsroom/blogs/random-samplings/2016/06/americas-age-profile-told-through-population-pyramids.html

Rosen, J. (2007). The People Formerly Known as the Audience. In Mandiberg, M. (Ed.) *The Social Media Reader* (13-16). New York, NY: New York University Press.

Ross, C., et al. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior 25*, 578-586.

Samek, A. (2016, September 23). The Association between Personality Traits and Voting Intentions in the 2016 Presidential Election [Blog post]. Retrieved from http://evidencebase.usc.edu/the-association-between-personality-traits-and-voting-intentions-in-the-2016-presidential-election/

Scott, C., et al. (2017). Time spent online: Latent profile analyses of emerging adults' social media use. *Computers in Human Behavior 75*, 311-319. Retrieved from https://dx.doi.org/10.1016/j.chb.2017.05.026

Shah, D.V., Cho, J., Eveland Jr., W.P., & Kwak, N. (2005, October 5). Information and Expression in a Digital Age: Modeling Internet Effects on Civic Participation. *Communication Research 32*(5), 531-565. Retrieved from https://dx.doi.org/10.1177/0093650205279209

Shaker, L. (2012, January 1). Local political knowledge and assessments of citizen competence. *Political Opinion Quarterly 76*(3), 525-537. Retrieved from https://doi.org/10.1093/poq/nfs018

Shearer, E. & Gottfried, J. (2017, September). News Use Across Social Media Platforms 2017. Retrieved from http://www.journalism.org/wp-content/uploads/sites/8/2017/09/PJ_17.08.23_socialMediaUpdate_FINAL.pdf

Sides, J., Tesler, M., & Vavreck, L. (2018). *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America.* Princeton, NJ: Princeton University Press.

Silverman, C. (2016, November 16). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. *BuzzFeed News*. Retrieved from https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook#.ei12Ya7XZ

Smith, A. & Anderson, M. (2018, March 1). Social Media Use in 2018. Retrieved from http://pewinternet.org/2018/03/01/social-media-use-in-2018

Smith, S. (2016, July 5). 5 facts about America's political independents. *Pew Research Center Fact Tank.* Retrieved from http://www.pewresearch.org/fact-tank/2016/07/05/5-facts-about-americas-political-independents/

Stata.com. (n.d.). Fractional outcome regression. Available at https://www.stata.com/features/overview/fractional-outcome-models/

Stolle, D. & Gidengil, E. (2010, March). What Do Women Really Know? A Gendered Analysis of Varieties of Political Knowledge. *Perspectives on Politics 8*(1), 93-109. Retrieved from https://doi.org/10.1017/S1537592709992684

Sunstein, C.R. (2007). *Republic.com 2.0.* Princeton, NJ: Princeton University Press.

Textor, J., van der Zander, B., Gilthorpe, M.K., Liskiewicz, M., & Ellison, G.T.H. (2016). Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology 45*(6), 1887-1894. Available at http://johannes-textor.name/papers/2017-ije.pdf. Tool available at http://www.dagitty.net/

Turccotte, J., York, C., Irving, J., Scholl, R.M., & Pingree, R.J. (2016, June 1). News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking. *Journal of Computer-Mediated Communication 20*(5), 520-535. Retrieved from https://doi.org/10.1111/jcc4.12127

United States Census Bureau. (n.d.) QuickFacts. Available at https://www.census.gov/quickfacts/fact/table/US#

Verba, S., Burns, N., & Schlozman, K.L. (1997, November). Knowing and Caring about Politics: Gender and Political Engagement. *The Journal of Politics 59*(4), 1051-1072. Retrieved from https://www.jstor.org/stable/2998592

Weeks, B.E., Ardèvol-Abreu, A., & Gil de Zúñiga, H. (2017, June 1). Online Influence? Social Media Use, Opinion Leadership, and Political Persuasion. *International Journal of Public Opinion Research 29*(2), 214-239. Retrieved from https://doi.org/10.1093/ijpor/edv050

Williams, R. (2018, April 10). Analyzing Proportions: Fractional Response and Zero One Inflated Beta Models [Handout]. Retrieved from https://www3.nd.edu/~rwilliam/stats3/FractionalResponseModels.pdf

WittyCookie. (2012, June 4). What are the major differences among Web 1.0, 2.0 and 3.0? [Blog post]. https://wittycookie.wordpress.com/2012/06/04/what-are-the-major-differences-among-web-1-0-2-0-and-3-0/

Woolley, J., Limperos, A.M., & Oliver, M.B. (2010). The 2008 Presidential Election, 2.0: A Content Analysis of User-Generated Political Facebook Groups. *Mass Communication and Society 13*, 631-652. Retrieved from https://doi.org/10.1080/15205436.2010.516864

Yong, A.G. & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology 9*(2), 79-94. Retrieved from https://www.doi.org/10.20982/tqmp.09.2.p079

YouGov. (2018). Terms and Conditions. Retrieved from https://today.yougov.com/about/terms-combined/#/privacy

Appendix A: Factor Analysis Output

Table 10: *Frequency Distribution of Social Media Use Response Patterns*

Pattern*	***Raw Count***	***Frequency (%)***
00000	27,192	42.09
00010	10,281	15.91
11111	4,517	6.99
00011	2,942	4.55
11110	1,920	2.97
01010	1,472	2.28
01000	1,316	2.04
11010	1,321	2.04
00110	1,268	1.96
00100	1,147	1.78
00001	1,066	1.65
10010	1,060	1.64
00111	1,035	1.60
10000	1,034	1.60
01011	984	1.52
11011	957	1.48
01110	724	1.12
01111	697	1.08
10110	660	1.02
10011	508	0.79
11000	502	0.78
10111	412	0.64
01001	269	0.42
11100	250	0.39
01100	227	0.35
10100	220	0.34
11001	200	0.31
10001	107	0.17
00101	105	0.16
11101	89	0.14
01101	70	0.11
10101	48	0.07
Total	**64,600**	**100.00**

Table 11: *Social Media Behavior Preliminary EFA Results A*

Variable	*Factor1*	*Factor2*	*Factor3*	*Factor4*	*Communality*
Post Content	0.7311	-0.2506	0.0311	0.0474	0.6006
Post Comment	0.7312	-0.1578	-0.1226	0.0229	0.5752
Forward Content	0.6602	-0.0220	0.1275	-0.0559	0.4557
Read/Watch Content	0.5756	0.2534	0.0118	0.0802	0.4021
Follow Event	0.5848	0.2860	-0.0411	-0.0465	0.4276

*Iterated Principal Factors Method, No Rotation, No Factor Limit

Table 12: *Social Media Behavior Preliminary EFA Results B*

Variable	*Factor1*	*Factor2*	*Communality*
Post Content	0.7433	-0.2791	0.6303
Post Comment	0.7189	-0.1249	0.5324
Forward Content	0.6532	-0.0161	0.4269
Read/Watch Content	0.5710	0.2315	0.3796
Follow Event	0.5880	0.2986	0.4349

*Iterated Principal Factors Method, No Rotation, 2 Factor Limit

Table 13: *Social Media Behavior EFA Results - No Omitted Loadings*

Variable	*ActiveEngagement*	*Consumption*	*Communality*
Post Content	0.8236	-0.0428	0.6303
Post Comment	0.6138	0.1527	0.5324
Forward Content	0.4351	0.2227	0.4269
Read/Watch Content	0.0710	0.5640	0.3796
Follow Event	-0.0028	0.6614	0.4349

*Iterated Principal Factors Method, Oblique Promax Rotation, 2 Factor Limit

Table 14: *Frequency Distribution of Political Knowledge Response Patterns*

Pattern*	***Frequency***	***Percent***	***Pattern****	***Frequency***	***Percent***
111111	22,123	35.08	100010	321	0.51
000000	7,482	11.86	001111	308	0.49
111110	3,028	4.8	101010	312	0.49
111100	2,870	4.55	011000	304	0.48
101111	2,808	4.45	110011	279	0.44
100000	1,532	2.43	010100	272	0.43
111101	1,525	2.42	110110	261	0.41
101100	1,220	1.93	000110	238	0.38
100100	986	1.56	001011	235	0.37
111011	965	1.53	100101	225	0.36
000011	912	1.45	111001	206	0.33
100111	884	1.4	100001	193	0.31
110111	873	1.38	101001	182	0.29
101000	864	1.37	010111	178	0.28
000100	848	1.34	110010	175	0.28
101011	806	1.28	001010	172	0.27
101110	711	1.13	011011	163	0.26
001000	648	1.03	001110	157	0.25
000010	607	0.96	110101	159	0.25
100011	600	0.95	011110	153	0.24
110100	568	0.9	000101	146	0.23
111000	558	0.88	010011	141	0.22
110000	484	0.77	110001	118	0.19
101101	445	0.71	010110	116	0.18
001100	443	0.7	011010	106	0.17
010000	431	0.68	001001	99	0.16
011111	424	0.67	001101	94	0.15
011100	381	0.6	011101	80	0.13
100110	376	0.6	010010	67	0.11
000111	373	0.59	011001	61	0.1
000001	347	0.55	010001	50	0.08
111010	325	0.52	010101	47	0.07
Subtotal	**57,447**	**91.07**	***Subtotal***	**5,618**	**8.91**
Total	**63,065 (99.98% of total dataset *n* – 64,600)**				

*From left to right: Party of Governor, Representative, Senator 1, Senator 2, Congressional Majority, Senate Majority. "1" indicates respondent answered correctly.
SOURCE: Recoding of variables from Ansolabehere and Schafner 2017

Table 15: *Political Knowledge Preliminary EFA Results A*

Variable	*Factor1*	*Factor2*	*Factor3*	*Factor4*	*Factor5*	*Communality*
House Majority	0.7071	0.3880	0.0004	0.0018	-0.0310	0.6515
Senate Majority	0.6855	0.3991	0.0060	-0.0060	0.0304	0.6302
Governor	0.7015	-0.1853	-0.1386	-0.0558	0.0148	0.5489
Representative	0.6773	-0.1909	0.1394	0.0582	0.0145	0.5182
Senator 1	0.7340	-0.2185	0.0696	-0.1056	-0.0147	0.6027
Senator 2	0.7057	-0.1818	-0.0747	0.1134	-0.0118	0.5497

*Iterated Principal Factors Method, No Rotation, No Factor Limit

Table 16: *Political Knowledge Preliminary EFA Results B*

Variable	*Factor1*	*Factor2*	*Communality*
House Majority	0.7096	0.3824	0.6498
Senate Majority	0.6879	0.3935	0.6280
Governor	0.6960	-0.1830	0.5178
Representative	0.6718	-0.1878	0.4866
Senator 1	0.7308	-0.2196	0.5823
Senator 2	0.7019	-0.1825	0.5260

*Iterated Principal Factors Method, No Rotation, 2 Factor Limit

Table 17: *Political Knowledge EFA Results – No Omitted Loadings*

Variable	*PersonalKnowledge*	*GeneralKnowedge*	*Communality*
House Majority	0.0624	0.7650	0.6498
Senate Majority	0.0354	0.7695	0.6280
Governor	0.6753	0.0668	0.5178
Representative	0.6642	0.0508	0.4866
Senator 1	0.7394	0.0365	0.5823
Senator 2	0.6789	0.0699	0.5260

*Iterated Principal Factors Method, Oblique Promax Rotation, 2 Factor Limit

Appendix B: Logit Regression Tables

Table 18: *Full Logit Regressions of Voting on Social Media, Mediated and Unmediated, with and without State Fixed Effects*

Model Number	1	2	3	4	5	6
State Fixed Effects		✓			✓	✓
Mediation Variables			✓	✓	✓	✓
Treatment of "Not Sure" Responses	N/A	N/A	Missing	Incorrect	Missing	Incorrect
Observations	47,400	47,400	35,120	46,359	35,120	46,359
Hosmer-Lemeshow Test p-Value	0.0363	0.2776	0.0021	0.1446	0.0048	0.2276
Social Media News:						
Consumption	1.135***	1.140***	1.118**	1.086*	1.119**	1.089*
Active Engagement	0.964	0.965	0.936	0.975	0.941	0.976
Consumption * Active Engagement	1.027	1.024	1.034	1.017	1.03	1.014
Political Knowledge:						
General			1.683***	1.356***	1.660***	1.348***
Personal			3.160***	2.684***	3.217***	2.740***
Other News Use	0.978	0.979	0.889	0.846***	0.891	0.847**
Age	1.029***	1.029***	1.024***	1.021***	1.024***	1.021***
Income	1.015**	1.018**	0.996	0.997	0.999	1.000
Female	1.05	1.055	1.124**	1.219***	1.131***	1.225***
Education ("Base = HS or Less"):						
Some college	1.402***	1.397***	1.223***	1.257***	1.223***	1.255***
2-year	1.378***	1.365***	1.208**	1.254***	1.197**	1.245***
4-year	1.255***	1.255***	1.035	1.060	1.033	1.061
Post-grad	1.272***	1.280***	1.049	1.059	1.05	1.063
Race (Base = "White"):						
Black	0.761***	0.802***	0.748***	0.774***	0.784***	0.809***
Hispanic	0.671***	0.754***	0.669***	0.697***	0.754***	0.786***
Asian	0.405***	0.426***	0.426***	0.415***	0.459***	0.441***
Native American	0.774	0.785	0.679*	0.833	0.700*	0.849
Middle Eastern	0.605	0.611	0.430	0.564	0.413	0.559
Mixed	0.873	0.894	0.885	0.872	0.921	0.894
Other	0.823	0.829	0.742	0.788	0.753	0.794
Party ID (Base = Independent):						
Democrat	1.112	1.112	1.186*	1.147*	1.187*	1.147*
Republican	1.149	1.165*	1.112	1.236**	1.141	1.248**
Other	1.642*	1.581*	1.669*	1.466	1.581*	1.411
Not Sure	0.509***	0.514***	0.566**	0.650**	0.571**	0.659**

Model Number	**1**	**2**	**3**	**4**	**5**	**6**
Ideology (Base = Moderate):						
Liberal	1.313***	1.304***	1.313**	1.229**	1.307**	1.223**
Conservative	1.135	1.136	1.09	1.099	1.085	1.101
Not Sure	0.411***	0.410***	0.562*	0.575***	0.576*	0.576***
Party ID # Ideology Interactions (Bases = Independent and Moderate):						
Democrat # Liberal	1.09	1.093	0.971	1.028	0.976	1.03
Democrat # Conservative	0.564***	0.570***	0.581***	0.619***	0.594***	0.623***
Democrat # Not Sure	2.685***	2.667***	2.377*	2.239***	2.314*	2.209***
Republican # Liberal	0.338***	0.349***	0.357***	0.380***	0.362***	0.392***
Republican # Conservative	1.141	1.132	1.089	1.032	1.077	1.026
Republican # Not Sure	2.471***	2.345***	4.273**	2.328**	3.759*	2.211**
Other # Liberal	0.705	0.761	0.615	0.727	0.687	0.784
Other # Conservative	1.047	1.069	0.879	0.988	0.912	1.012
Other # Not Sure	0.585	0.614	0.593	0.493*	0.617	0.508
Not Sure # Liberal	0.819	0.826	1.005	0.835	1.039	0.849
Not Sure # Conservative	0.802	0.787	1.363	0.909	1.358	0.893
Not Sure # Not Sure	1.338	1.33	1.388	1.077	1.356	1.068
State (Base = CA):						
Alabama		0.991			1.016	1.062
Alaska		1.819			1.799	1.810
Arizona		0.991			0.999	1.025
Arkansas		1.150			1.448	1.251
Colorado		1.308*			1.406*	1.415**
Connecticut		1.044			1.095	1.104
Delaware		1.085			1.234	1.203
Florida		0.875			0.906	0.919
Georgia		1.016			1.176	1.111
Hawaii		1.094			0.745	0.959
Idaho		0.897			0.901	0.844
Illinois		1.073			1.125	1.134
Indiana		1.400**			1.390*	1.520***
Iowa		1.672***			1.794**	1.549**
Kansas		1.248			1.431	1.240
Kentucky		0.867			0.890	0.859
Louisiana		0.966			0.871	1.001
Maine		1.904**			2.248**	1.832**
Maryland		1.207			1.360*	1.325*
Massachusetts		0.958			0.961	1.010
Michigan		1.216*			1.321*	1.300**
Minnesota		1.308*			1.348*	1.282*

Model Number	1	2	3	4	5	6
Mississippi		0.923			0.806	0.899
Missouri		1.381**			1.487**	1.368**
Montana		1.462			1.442	1.440
Nebraska		1.140			1.083	1.119
Nevada		1.452*			1.288	1.647**
New Hampshire		1.327			1.456	1.230
New Jersey		0.828			0.891	0.860
New Mexico		1.230			1.240	1.264
New York		0.908			0.924	0.973
North Carolina		1.259*			1.260*	1.315**
North Dakota		1.360			0.943	1.371
Ohio		1.326**			1.316*	1.365**
Oklahoma		1.086			0.877	1.134
Oregon		1.519***			1.649***	1.598***
Pennsylvania		1.309**			1.332**	1.402***
Rhode Island		0.851			1.101	0.811
South Carolina		1.169			1.282	1.178
South Dakota		0.963			0.914	0.904
Tennessee		0.900			0.940	0.9200
Texas		0.853			0.886	0.878
Utah		1.197			1.250	1.103
Vermont		1.033			1.164	1.097
Virginia		1.007			0.969	1.055
Washington		1.373**			1.425*	1.410**
West Virginia		0.937			1.013	1.063
Wisconsin		1.222			1.265	1.159
Wyoming		0.875			1.105	0.851

Exponentiated coefficients, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: *Fractional Logit Regressions of Political Knowledge on Social Media News Use*

(Regression #) Political Knowledge Outcome:	(7) General	(8) Personal
Social Media News:		
Consumption	1.243***	1.187**
Active Engagement	0.818***	0.821***
Consumption * Active Engagement	1.121***	1.143***
Other News Use	1.103	1.057
Age	1.027***	1.030***
Income	1.041***	1.050***
Female	0.665***	0.761***
Education (“Base = No HS”)		
High school graduate	0.97	1.044
Some college	1.459**	1.568***
2-year	1.284	1.326*
4-year	1.512***	1.720***
Post-grad	1.705***	1.867***
Race (Base = “White”)		
Black	0.807**	0.760***
Hispanic	0.707***	0.818*
Asian	0.744**	0.722*
Native American	0.78	0.697
Middle Eastern	2.974**	1.197
Mixed	1.239	1.199
Other	1.148	1.939**
Party ID (Base = Independent)		
Strong Democrat	0.631***	0.688**
Democrat	0.993	0.915
Republican	1.094	0.843
Strong Republican	0.584***	0.537***
Not Sure	1.33	1.09
Ideology (Base = Moderate)		
Very Liberal	1.596	1.014
Liberal	0.867	0.636*
Conservative	1.251	0.76
Very Conservative	2.315**	1.257
Not Sure	0.640*	0.334***
Party ID * Ideology Interactions (Bases = Independent & Moderate)		
Strong Democrat * Very Liberal	1.445	1.831
Strong Democrat * Liberal	2.298***	2.899***
Strong Democrat * Conservative	0.468***	0.759
Strong Democrat * Very Conservative	0.189***	0.389*
Strong Democrat * Not Sure	1.105	2.146*
Democrat * Very Liberal	1.555	2.801**
Democrat * Liberal	1.783**	1.981**
Democrat * Conservative	0.371***	0.578*
Democrat * Very Conservative	0.127***	0.261**
Democrat * Not Sure	1.154	2.209

Table 10 (Cont.): *Frac. Logit Regressions of Political Knowledge on Social Media Use*

(Regression #) Political Knowledge Outcome:	(3) General	(4) Personal
Republican * Very Liberal	0.73	0.366*
Republican * Liberal	0.546	0.748
Republican * Conservative	0.98	1.750**
Republican * Very Conservative	0.796	2.380*
Republican * Not Sure	2.073	0.94
Strong Republican * Very Liberal	0.727	0.612
Strong Republican * Liberal	1.257	0.763
Strong Republican * Conservative	2.096***	3.100***
Strong Republican * Very Conservative	1.18	1.796
Strong Republican * Not Sure	1.21	3.378*
Not Sure * Very Liberal	0.0484**	0.739
Not Sure * Liberal	0.676	0.341
Not Sure * Conservative	0.283*	0.494
Not Sure * Very Conservative	0.141***	0.0979***
Not Sure * Not Sure	0.487	1.257
Observations	35105	35105

Exponentiated coefficients, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$