Erasmus School of Economics

ERASMUS UNIVERSITY ROTTERDAM

MASTER THESIS

BEHAVIORAL ECONOMICS

Do you plan to vote? Did you vote?

The Gap between Promise and Practice in the Age of Social Media

Jackson Kent

637817

Supervisor: Dr. Georg Granic

Second Assessor: Dr. Dana Sisak

AUGUST 2023

Table of Contents

Abstract

Why do some eligible voters in elections say they will vote but never follow through in doing so? This thesis sheds light on this phenomenon of "over-reporting" through the lens of political social media engagement during the 2020 US presidential election. Using survey data from the Cooperative Election Study (CES) in tandem with data on validated voters, an empirical framework is implemented utilizing regression analysis to examine the associations between political activity on social media and one's propensity to over-report their voting intentions. This study finds that politically active members of social media are associated with lower likelihoods of over-reporting when compared to offline individuals. Additionally, further analysis revealed that identifying as conservative is more heavily associated with over-reporting behavior than identifying as liberal. These findings are relevant to policy makers and campaign organizers seeking to harness digital platforms more effectively, with an emphasis on gauging which demographics are more likely to misrepresent their voting intentions. While this thesis takes strides in unearthing the relationship between political social media engagement and over-reporting, further research is necessary to disentangle the underlying mechanisms of over-reporting behavior.

KEYWORDS: Over-reporting - Social Desirability Bias - Political Elections - Social Media

Introduction

The bedrock of any democratic society is the fair and accurate representation of the public's sentiment, often gauged through polling data. However, polling discrepancies in recent elections have underscored the challenges in gathering a true representation of the public's will (Kennedy et al, Pew Research 2021). A critical yet underexamined factor that contributes to this issue is over-reporting - the phenomenon where individuals claim to intend to or have voted, but did not actually vote. Over-reporting has been identified as a significant source of bias in election polling since as early as 1942 (Cantril and Harding, 1942), but was brought to further attention in 1968 (Clausen, 1968) when methods for voter validation revealed a discrepancy between the tally of survey respondents who claimed to have voted and the actual validated count. This research was owed to the introduction of methods for validating voting records on an individual level, allowing not only for accurate measurements of who turns out to vote, but also a novel approach in studying voter turnout. These findings have been found repeatedly into the 21st century, and as such researchers have continued to grapple with various explanatory theories for potential causes (Silver, Anderson, and Abramson, 1986; Ansolabehere and Hersh, 2012; Karp and Brockington, 2005; Brenner and DeLamater, 2016).

The lexicon of prior literature on this topic has found social desirability bias to be a salient explanation for this phenomenon, a behavioral tendency where survey respondents are drawn towards answers they perceive as more socially acceptable (Clausen, 1968; Silver, Anderson, and Abramson, 1986; Ansolabehere and Hersh, 2012). Much of this past literature has focused on using econometric methods to empirically identify the characteristics of over-reporters, finding that they typically belong to groups where political participation is deemed as socially desirable: the highly educated, elders, church attendees, partisans, etc. (Ansolabehere and Hersh, 2012). It is suggested that due to their societal roles, individuals from these groups would feel greater shame for not participating in voting, thereby making them more likely to misrepresent their voting behavior. The intent of this thesis is to further empirical exploration into relevant demographic traits that may also foster this sort of social desirability bias, ultimately adding to what we understand about the nature of over-reporting. By learning which demographic traits or behavioral patterns are indicative of over-reported vote intentions, polls can correct for this bias in their estimates mid-election season and make more informed predictions on who will turn out to vote.

As the digital landscape of today's world evolves, the terrain of social desirability bias and over-reporting has dramatically changed. Heightened political engagement on social media during the 2020 US presidential election offers a novel landscape to examine these phenomena in tandem. Social media platforms like Twitter, Facebook, or Instagram, with feeds teeming with politically charged posts during this time, can exert considerable social pressure on users to align with the dominant narrative of their respective network. Consequently, this may compel individuals to over-report their political participation to conform to perceived social expectations. As such, this thesis uses the context of the 2020 US election with social desirability bias as a theoretical framework to answer the research question: How does political engagement on social media affect over-reporting behavior? Leveraging data from the Cooperative Election Study (CES) survey in conjunction with validated voting records, this thesis provides an empirical framework to investigate the behavior of political social media users when it comes to following through on their voting intentions. This thesis identifies two avenues wherein social desirability bias is theorized to have permeated the heart of American politics on social media during the 2020 presidential election season: the consequences of an unprecedented surge in political activity, and the polarized ideological landscape.

Amidst the societal divides brought about by Covid-19, the BLM Movement, and the election itself, social media became a hotbed of political activity and discourse, with those less involved risking being seen as outliers (Miller et al., 2022; Drakulich and Denver, 2022). For example, users with limited political engagement, who may have merely viewed a politically charged post from a friend without having posted themselves, may feel compelled by the overwhelmingly politicized environment to express a voting intention they do not truly hold. Being politically inactive in an environment where taking an angle on politics is the social norm may lead individuals to misrepresent their political participation when put on the spot. Additionally, this thesis delves into how social desirability bias might influence social media users across the political spectrum. In a climate where adopting a political stance is almost mandatory, the pressure to conform can make individuals overstate their political involvement. This exploration is particularly relevant given the intense partisan divides characterizing recent elections, none of which have witnessed the high level of social media prevalence as was seen in the 2020 US election.

This study found that among nonvoters, those engaged in political activities on social media were associated with lower likelihoods of over-report when compared to those offline. A notable yet slight negative correlation was observed between increased political social media activity and over-reporting, but this was evident only when factoring in ideology. While a pronounced positive association between conservative ideologies and over-reporting emerged, it's crucial to note that this relationship was independent of any interaction with social media engagement. In other words, the role of social media was not a determining factor in the link between over-reporting and ideological leaning. Despite providing little evidence for or even contradicting many of the hypotheses, these results contribute to a deeper understanding of the relationship between political polarization, social media usage, and over-reporting. It suggests that social desirability bias, while universal, may have differential effects based on one’s ideological leanings and their decision to be politically engaged on social media or not. These findings carry significant implications for the design and interpretation of future polling surveys. In terms of design, surveys may need to incorporate questions that more accurately gauge respondents' social media usage and political engagement online beyond the more basic questions asked on the CES. Furthermore, when interpreting results, analysts should take into account the potential bias of over-reporting, particularly among certain demographics of social media users or ideological groups. This becomes increasingly important as our society continues to grow more digitized and/or politically polarized. Further research would provide a more comprehensive understanding of this intricate dynamic and better inform future polling methodologies.

Theory Review

As is explained in the introduction, the reliability of any democratic process is rooted in its ability to accurately measure public sentiment, with many recent elections having failed in this respect. While the polling discrepancies of recent elections were largely a product of statistical malfeasance in calculating the likelihoods of voter turnout among varying demographics (Kahu, 2022), there remains the inherent gap in self-reported survey data to fill. A wealth of research on over-reporting has taken place over the past half-century, with the near entirety of it focusing on the post-hoc perspective where subjects are surveyed *after* the election and are asked to answer whether they *had* voted or not. While this approach does well to inform on how voter turnout numbers may have been biased in an election after the fact, it does little good for campaign managers who need information on how surveys can be biased during election season when strategies can still be adjusted. This study deviates from the traditional post-election analysis of over-reporting, offering a novel, pre-hoc perspective. By analyzing how survey respondents intend to vote when contrasted with how they vote in reality, a unique angle on the over-reporting problem is observed that helps to explain part of the turnout problem that election polling currently does not address.

Past research has utilized various experimental methods to discern the behavioral traits responsible for over-reporting, with many pointing to social desirability bias as a primary culprit. To state briefly, social desirability bias is defined as the inclination for survey respondents to choose the "socially desirable" answer on surveys, particularly on taboo topics (Grimm, 2010). Findings from Hanmer et al. (2013) suggests that individuals can be tempted to portray themselves as active voters in elections when they in fact are not, unless they know they can be caught lying. Using the "bogus-pipeline technique," the researchers told respondents that their answers could be verified, and would be found out if caught lying. The presence of social desirability bias was inferred when it was common knowledge between the surveyor and the respondent that their answer could be verified, reducing the pressure to lie about engaging in the socially desirable behavior, in this case voting. An experiment by Duff et al. (2007) provided further evidence for the presence of social desirability bias, demonstrating that providing socially acceptable excuses for not voting, or "face-saving questions" within the survey can reduce over-reporting in the American National Election Study telephone survey. Merely changing the wording from "I did not vote" to "I meant to vote, but did not have time this year" lowered over-reporting frequencies. While these methods yielded significant effect sizes, the studies were limited to the American population and as such lack generalizability. Morine-Chasse et al. (2017) sought to overcome this shortcoming by also implementing the "face-saving questions" method, but this time over a panel of 19 online surveys across European and Canadian countries. They found that by slightly altering the questioning to signify social acceptance, they could decrease over-reporting by a mean average of 7.6 percentage points across all countries surveyed. The relevance of these findings to this study are two-fold: the presence of social desirability bias is clear, and social desirability bias is also present (to varying degrees) in settings with less immediate external pressure from interviewers, such as online or over-the-phone surveys. A study conducted by Belli et al. (1999) suggests that over-reporting can be a combination of social desirability bias and memory failure, finding that when subjects forget whether they had voted in the election, they err towards the socially desirable response that they did in fact vote.

Research utilizing self-reported survey data in tandem with validated vote data have also revealed empirical evidence in support of the social desirability theory hypothesis. The goal of much of this research was to identify key independent variables among nonvoters that capture social desirability and subsequently explain over-reporting behavior. It was assumed in earlier research that the population of over-reporters had unique characteristics distinct from regular voters that decisively set them apart. However, more recent studies utilizing advanced validation techniques have consistently found that a core set of voter-like characteristics can explain most over-reports. Notably, over-reporters tend to look like actual voters, and are typically *older, politically engaged, partisan, highly educated*, and *church-going* people (Silver, Anderson, and Abramson, 1986; Ansolabehere and Hersh, 2012). The dominant theory that explains over-report behavior among these nonvoters that look like voters is the social desirability to appear as a member of the voting group, despite contradictory behavior. Jackman and Muha (1984) argue that middle to upper class voters understand that voting is a socially desirable behavior among their peers, and feel the need to express their conformity to the status-quo in this way. Hence, they tend to misrepresent their voting behavior more often than those whose peers are less likely to be engaged in the democratic process.

To further investigate the issue of social desirability and over-reporting, this thesis turns to the distinctive political landscape of the 2020 US presidential election characterized by the unprecedented prevalence of social media. The unique circumstances surrounding the election, including widespread lockdown measures imposed due to the COVID-19 pandemic, dramatically amplified the reliance on digital platforms for communication, campaigning, and the dissemination of political information (Bail et al., 2020). Sources from data-collection company Socialbakers indicate that 79% of eligible voters in the US actively used social media. As a result, many of these individuals took to social media platforms to express their political beliefs and engage in conversations around political topics. According to a 2020 Statista study, Twitter saw an increase of 78% increase in user engagement compared to 2018, coinciding with increased political activity surrounding the election. The acute rise of universal social media usage - especially in the context of recent US presidential elections - provides a novel angle on social desirability bias to examine in relation to over-reporting, leading to the main research question of this study:

What are the effects of political social media usage on over-reporting behavior?

Social desirability is amplified on social media platforms where individual’s posts and interactions are visible to a broader audience where networks are much vaster than in reality, yielding more opportunity for social scrutiny. In an increasingly connected world, social media platforms like Facebook, Twitter, and Instagram have not only become hubs for interpersonal communication, but also influential spaces for political discourse and activism. These platforms allow for instantaneous sharing of political opinions, news, and other election-related content, catalyzing political engagement. However, the public nature of these platforms can create an environment where users feel pressured to present themselves in a certain light to be socially accepted. For instance, posting about voting in an election can be seen as a socially desirable action, and as a result, users who did not vote might still post about voting to match societal expectations or maintain a certain image among their peers (Krämer & Winter, 2008). Bernstein et al (2001) summarize the essence of the over-reporting issue in relation to social desirability in this way: "*those who are under the most pressure to vote are the ones most likely to misrepresent their behavior when they fail to do so*." This theory falls in line with research from Ansolabehere and Hersh (2012) who found nonvoters who over-report tend to look very similar to voters, as they want to associate with what their group deems as socially desirable - it can be very easy to look like someone you are not on social media. As such, social media platforms are a likely place for social desirability and hence the over-reporting issue to manifest itself. The first research hypothesis follows:

H1: Political activity on social media is positively associated with over-reporting vote intentions when compared to those who are not

The community of those who are politically engaged on social media is far from monolithic. Of the demographic clusters of users across various platforms, this thesis identifies active and passive users as distinct groups that may exhibit different socially desirable behaviors. Active users are those who interact with political content in the public sphere, whether it be posting a photo or video, sharing the content of others, or commenting on the posts of others. Passive users are defined as those who choose to consume media in their own private space, rarely letting their presence or inclinations be known to others via more social avenues of interaction. In the prior literature on over-reporting, researchers have pointed to "internal" vs. "external" pressures that people face in relation to misrepresentation of voting behavior. Hanmer et al. (2013) argues that the over-reporting issue is a matter of internal pressure rather than external, as people often seek to reinforce their own self-image. On the other hand, it is found that external pressure, i.e. pressure that is felt from those in one's immediate vicinity is less explanatory of over-reporting. Hanmer cites Ansolabehere and Hersh (2012) in their finding that in the 2008 Congressional Election Study survey data, 50% of validated nonvoters said that they had voted, indicating a lack of external pressure due to the anonymous nature of the online survey, and the presence of internal pressure.

Despite what researchers may claim about the fully anonymous nature of their own studies, the assumption of full anonymity in online surveys may not be entirely convincing to those individuals taking the survey who cannot verify the true anonymity of their answers. Although researchers collect data anonymously, panel participants often reveal identifying information to the panel companies who administer the survey (in the case of the CES survey, YouGov). This may raise doubts for participants about the extent of personal information available to the researcher, possibly inducing some level of external pressure. Despite the interplay between internal and external pressures in online surveys being more complex than previously assumed, this thesis follows the theoretical route taken by Bernstein et al. (2001) in their intuition that internal pressure felt in the form of self-induced guilt is more explanatory of over-reporting behavior than potential external pressures. They reject the status quo that over-reporting results from the pressure to appear desirable in front of interviewers (or researchers), rather arguing that it stems from the guilt we feel when we disappoint ourselves by not following through with our self-perceived obligations. Additionally, the increased pressure from the societal groups that these individuals belong to or identify with (i.e., social media) creates added guilt for not living up to their own expectations of how they should be (Bernstein et al, 2001).

Since active social media users are typically the center of public attention, they are more likely to feel external pressure from peers to vote. Passive users, on the other hand, are more likely to feel internal pressure to be like their more active peers, ultimately seeking self-confirmation by (untruthfully) painting themselves as a member of the status-quo voting community. As such, the subsequent hypothesis follows:

H2: Passive social media users are more likely to over-report their voting intentions than their more active counterparts

In addition to the active versus passive user divide, the heightened partisan divide on social media during the 2020 election season presents grounds for fresh perspective on the over-reporting issue. The highly polarized political landscape led to unprecedented levels of political activity on social media, which may have fueled social desirability bias and hence over-reporting across party lines. The partisan divide between Democrats and Republicans reached peak levels during this election, especially due to pressing controversial issues including Covid-19, the Black Lives Matter movement, and electoral integrity (Drakulich and Denver, 2022). The hyperpolarization of these issues was compounded by the emergence of "echo chambers" on social media, where extreme views on both sides of the political spectrum were amplified yet contained in their respective groups (Barbera, 2020). The stark nature of this divide warrants further investigation into how users between the two parties may have been affected by the strong societal pressure to align to one group or another, potentially causing misrepresentation of voting intentions. Therefore, this thesis hypothesizes:

H3: Due to the vastness in political opinions between liberals and conservatives in 2020, there will be significant variance in over-reporting behavior between political ideology affiliations

Methods

Data

This study is made possible by data from the Cooperative Election Study (CES) survey conducted during the year of the 2020 US election (Ansolabehere, Schaffner, and Luks, 2021. The CES is an extensive, nationally representative survey that collects data on various aspects of voter behavior and demographics, holding observations on 61,000 voting-age Americans who took the survey online. As it pertains to this study, the CES survey contains data on social media usage, pre-election voting intentions, and post-election validated votes for each respondent. It also contains a wealth of demographic information per observation, including ideology and party preference. The comprehensive nature of the CES survey data combined with the unique circumstances of the 2020 US election provide this study with a unique opportunity to examine over-reporting behaviors in respect to social media usage.

Variable and Model Estimation

This section specifies the formulation choices made for each variable used in the respective regression analyses, which have been carefully designed to investigate their corresponding hypotheses and ultimately address the overarching research question.

The dependent variable of interest in this study is the *Over-Reported Vote*, which is specified as respondents who indicated intent to vote but were ultimately validated as nonvoters. The indicator is created from variables “CC20_363” and “CL_2020gvm,” which are the stated intentions to vote and validated votes, respectively. The variable is coded as zero for nonvoters who indicated that they would not vote, and one for those of whom who indicated they intend vote. The estimation technique for each of the statistical test used in this study follows from the binary nature of the dependent variable, and as such the choice model for each test is a standard logistic regression.

The following independent variables used in this thesis are adapted from Ansolabehere and Hersh (2012), who confirm in their own OLS and logistic regression models these demographic features to be largely determinant of over-reporting. Other prior literature listed in this paper, particularly Bernstein et al. 2001 and Silver, Anderson, and Abramson, 1986, also accept these variables as primary determinants of over-reporting; this thesis follows in line and controls for their effects in each model. Despite many of the following variables having discrete categories in their mode of specification, all those that are non-binary, aside from *Age*, are included in models from the aforementioned authors as continuous variables. The hierarchal nature of each variable is the primary consideration in justification of their implementation as continuous, as each variable ranks “least” to “most” of their respective categories, i.e. “least educated” to “most educated.” This study follows the trail led by past researchers and utilizes these variables as such, while also exploring the robustness that an inclusion of the variables as categorical may have to offer. Therefore, this study analyzes the continuous specification of each model, as done in prior studies, in tandem with their categorical counterpart where each variable is input as categorical.

Education and *Income* are derived from values ranging from “No HS” and “0-$10,000” to “Postgraduate” and “$100,000+,” respectively, and are re-coded into five-category indicators. The categories for *Education* are “No Highschool,” “Highschool Graduate,” “Some College,” “College Degree,” and “Postgraduate Degree.” Categories for *Income* are “$0-29,999,” “$30,000-49,999,” “$50,000-99,999,” and “$100,000+,” and “Prefer not to say.” Categories for *Church attendance* is also a range of values that is condensed into just four categories for this analysis, and is ordered: “Never,” “Few times a year,” “Few times a month,” “At least once a week,” and “Don’t know.”

Black and *Other non-White* are re-coded from the “race” variable and are each turned into binary indicators. The *Married* and *Female* variables each receive a similar treatment, where they take the value of “0” for all values that are not married or not female, and “1” for married or female.

Political Interest is gauged from the question titled “newsint” on the survey that asks participants how closely they pay attention to politics in the news and is condensed to a scale between zero and three, ranging from “Hardly at all,” to “Most of the time.” *Ideological strength* is a variable that measures how devoted the respondent is to their political ideology, also ranging zero to three as “Weak,” “Moderate,” and “Strong.”

Age is specified as five categories: 18-24, 25-34, 35-44, 45-54, and 54 and over.

In addition to these independent variables, other survey questions are utilized from the same CES dataset to introduce new variable specifications that this thesis hypothesizes to also be determinant of over-reporting behavior.

Is Politically Engaged is a binary indicator that either places subjects into a group that has been politically engaged on social media in the past 24 hours, or one that has not. An observation is coded positively as 1 if they have engaged in any of the possible political social media activities over the past 24 hours, and 0 if they participated in none. The five possible actions are:

- posted a story, video, or link about politics,
- forwarded a story, video, or link about politics to a friend,

- posted a comment about politics,
- read a story or watched a video about politics,
- or followed a political event.

This variable serves as the main independent variable of interest in respect to H1, and is expected to have a positive coefficient, as to say that those who are politically engaged on social media are more likely to over-report. The model is as follows:

$$\begin{aligned}
\mathit{Model}\ 1\text{: } \Pr(\mathit{Over\ Report} = 1 \mid \beta x_1,\ \beta x_2, \dots, \beta x_n) \\
= \beta 0 + \beta 1 \mathit{Is\ Politically\ Engaged} + \beta 2 \mathit{Education} + \beta 3 \mathit{Income} \\
+ \beta 4 \mathit{Black} + \beta 5 \mathit{Other\ non\ White} + \beta 6 \mathit{Married} + \beta 7 \mathit{Church\ Attendance} \\
+ \beta 8\ \mathit{Age} + \beta 9 \mathit{Female} + \beta 10 \mathit{Ideological\ Strength} \\
+ \beta 11\ \mathit{Political\ Interest} + \epsilon
\end{aligned}$$

Once the effects of being politically engaged on social media are understood, the sample is further restricted to nonvoters who confirmed their engagement on social media within the past day. Derived from this is the *Social media engagement* variable adopted from Piatak and Mikkelson (2021), and aims to understand how escalating social media usage influences over-reporting among these engaged individuals. In their research, the social media engagement variable is a quasi-continuous additive measure composed of a selection of possible actions that the survey respondent indicated they had partaken in on social media in the past 24 hours, as is listed in bullet points above. Full engagement across all five activities yields a score of "5," while the least engagement results in a score of "0."

While there are just six potential categories for this variable, it is treated as continuous in Piatak and Mikkelson (2021)'s analysis, with the purpose of examining the effect of increasing social media engagement on over-reported vote intentions. Treating this variable as continuous is logical in this sense, as its categorical counterpart does not capture the nuance of each individual social media action. As a higher score translates to higher social media activity, the specification of this variable as continuous is warranted here. Additionally, it may account for micro variation in political social media engagement that would otherwise remain latent in the model, offering flexibility beyond the rigid categories. For example, sharing a political story within a large group chat of friends online might lie between the acts of merely forwarding and publicly posting it. The inclusion of this variable will shed light on how over-reporting behavior varies between types of users; those with lower scores can be considered more passive users, whereas those with higher levels of engagement are considered more active users. As such, the following model will test the prediction of H2 that the likelihood of over-reporting decreases as levels of political social media engagement increase, with the model specifying as follows:

$$\begin{aligned} \mathit{Model}\ 2\text{: } \Pr(\mathit{Over\ Report} = 1 \mid \beta x_1,\ \ \beta x_2, \dots, \beta x_n) \\ = \beta 0 + \beta 1 \mathit{Social\ Media\ Engagement} + \beta 2 \mathit{Education} + \beta 3 \mathit{Income} \\ + \beta 4 \mathit{Black} + \beta 5 \mathit{Other\ non\ White} + \beta 6 \mathit{Married} + \beta 7 \mathit{Church\ Attendance} \\ + \beta 8\ \mathit{Age} + \beta 9 \mathit{Female} + \beta 10 \mathit{Ideological\ Strength} \\ + \beta 11\ \mathit{Political\ Interest} + \epsilon \end{aligned}$$

Passive user is created as an alternative specification to test H2 and divides the set of possible political engagement into two groups, with one group engaging in solely passive activities. A "passive" user is coded as 1 if they have seen a political post or followed a political page, while simultaneously having engaged in none of the following activities: posted a story, video, or link about politics - forwarded a story, video, or link about politics to a friend - or posted a comment about politics. Having seen a political post or following a political page are considered the most passive forms of usage, with the other three forms of engagement being considered active. This thesis hypothesizes that passive users will be more likely to over-report when compared to non-passive users, and will be tested using the following model:

$$\begin{aligned} \mathit{Model}\ 3\text{: } \Pr(\mathit{Over\ Report} = 1 \mid \beta x_1,\ \ \beta x_2, \dots, \beta x_n) \\ = \beta 0 + \beta 1 \mathit{Passive\ User} + \beta 2 \mathit{Education} + \beta 3 \mathit{Income} + \beta 4 \mathit{Black} \\ + \beta 5 \mathit{Other\ non\ White} + \beta 6 \mathit{Married} + \beta 7 \mathit{Church\ Attendance} + \beta 8\ \mathit{Age} \\ + \beta 9 \mathit{Female} + \beta 10 \mathit{Ideological\ Strength} + \beta 11\ \mathit{Political\ Interest} + \epsilon \end{aligned}$$

As the examination of active versus passive users on social media is novel to this field, analysis utilizing an additional variable, *User Engagement Category*, adds an additional layer of robustness to findings from the previous two models. *User Engagement Category* is another division of the types of possible social media engagement actions that splits users into two groups based on a "user engagement score." The score is an additive measure of each activity, where the more "active" activities are more heavily weighted. The variable is as follows:

$$\textit{User Engagement Score} = \textit{Saw Story} + \textit{Followed Political Page} * 2 + \textit{Forwarded story} * 3 + \textit{Posted Comment} * 3 + \textit{Posted Story} * 4$$

After the score is calculated for each observation, *User Engagement* Categories is specified by placing the top 50th percentile into the "active" category and is coded as 0, with the bottom 50th being allocated to the "passive" category and is coded as 1. Following from the previous model, it is hypothesized that being a "passive" user will have a positive effect on the likelihood of over-reporting when compared to "active" users, yielding the following model:

$$\begin{aligned} \textit{Model } 4\text{: } \Pr(\textit{Over Report} = 1 \mid \beta x_1,\ \beta x_2, \dots, \beta x_n) \\ = \beta 0 + \beta 1 \textit{User Engagement Category} + \beta 2 \textit{Education} + \beta 3 \textit{Income} \\ + \beta 4 \textit{Black} + \beta 5 \textit{Other non White} + \beta 6 \textit{Married} + \beta 7 \textit{Church Attendance} \\ + \beta 8\ \textit{Age} + \beta 9 \textit{Female} + \beta 10 \textit{Ideological Strength} \\ + \beta 11\ \textit{Political Interest} + \epsilon \end{aligned}$$

To test the final hypothesis, the *Social Media Engagement* measure is interacted with *Ideology* to determine the effects of varying levels of social media usage in conjunction with ideology on over-reporting behavior. *Ideology* is a categorical measure that ranges from "Very Liberal," "Liberal," "Moderate," "Conservative," "Very Conservative," and finally "Not Sure." By examining how social media usage affects each of these categories independently, support for or against H3 can be found.

$$\begin{aligned} \textit{Model } 5\text{: } \Pr(\textit{Over Report} = 1 \mid \beta x_1,\ \beta x_2, \dots, \beta x_n) \\ = \beta 0 + \beta 1 \textit{Social Media Engagement} * \textit{Ideology} + \beta 2 \textit{Education} \\ + \beta 3 \textit{Income} + \beta 4 \textit{Black} + \beta 5 \textit{Other non White} + \beta 6 \textit{Married} \\ + \beta 7 \textit{Church Attendance} + \beta 8\ \textit{Age} + \beta 9 \textit{Female} \\ + \beta 10 \textit{Ideological Strength} + \beta 11\ \textit{Political Interest} + \epsilon \end{aligned}$$

Additionally, the authors of the CES study provide "common weights" they suggest applying as an argument in the model command to correct for possible sampling error. Since the survey is designed to oversample certain populations, the weights serve to make the data representative of the true population from which the sample was drawn, with the aim of making the findings as generalizable as possible. However, the same authors in their own research on over-reporting, using this same CES dataset, note that they do not employ the use of weights on their logistic regression models, as they would a linear OLS regression model (Ansolabehere and Hersh, 2012). Given that this thesis solely utilizes the logistic regression, it follows suit in not employing weights. Descriptive statistics of the independent variables of interest are depicted below in Table 1. Descriptive statistics of the additional variables controlled for throughout this analysis can be found in Table A.1 of the Appendix.

Table 1

Descriptive Statistics of Nonvoters with respect to Independent Variables of Interest

	Frequency	Frequency (%)	Cum. Frequency (%)
Is Politically Engaged			
Yes	2941	70.29	70.29
No	1243	29.71	100
SME Score			
0	1243	29.71	29.71
1	1031	24.64	54.35
2	763	18.24	72.58
3	532	12.72	85.3
4	532	8.41	93.71
5	263	6.29	100
Passive User			
Yes	1080	25.81	25.81
No	3104	74.19	100
User Engagement Cat.			
Passive	2090	49.95	49.95
Active	2094	50.05	100
Ideology			
Very Liberal	615	14.7	14.7
Liberal	750	17.93	32.62
Moderate	1210	28.92	61.54
Conservative	897	21.43	82.98
Very Conservative	572	13.67	96.65
Not Sure	140	3.35	100

Note: SME = Social Media Engagement

Results

While the data contains observations from 61,000 respondents, the sample shrinks to 39,198 observations when restricted to observations with no NA values across all variables of interest. Additionally, since this study takes interest in the proportion of over-reporters within the sample of all nonvoters as opposed to the full population (Ansolabehere and Hersh, 2012; Abramson, Anderson and Silver, 1986), the data is further restricted to just 4,167 observations when purely considering nonvoters with no NA values in the main variables included for hypothesis testing. When looking at only those engaged on social media, the sample decreases further to 2,985 respondents.

Of note in the descriptive statistics, the majority of nonvoters appear to be active on social media to at least some capacity. However, the distribution of social media engagement types is skewed toward the passive end, with many users having only engaged in one or two actions. The statistics for *Passive User* being inconsistent with this only show that many users do not meet the specific criteria for being a passive user as specified, rather than revealing a mostly active sample of users. *User Engagement* shows a more robust depiction of active vs passive users, with almost identical counts in each group. Lastly, it is shown that *Ideology* is most densely concentrated in the "Moderate" category with each ideological group shrinking on each side as they approach further partisanship.

Before looking into regression results, preliminary data exploration is telling as to how the independent variables in question relate to over-reporting. Table 2 shows the proportion of each sample that over-reports by category. The first column mirrors the descriptive statistics table, displaying the frequency of each category. The next column restricts that frequency to those who intended to vote, representing the over-reporting nonvoters. The final column reveals the proportion of over-reporters in relation to the total nonvoters in each category. Immediately it is observed that the vast proportion of nonvoters had intended to vote, with almost every category surpassing an 80% composition of over-reporters.

Looking more closely at each variable, it is seen in the first two rows that being politically engaged online is hardly different than not being engaged in terms of the likelihood of over-reporting, foreshadowing a lack of evidence for H1. When looking at variables used for testing H2, the Social Media Engagement scores give little evidence showing that lower social media engagement levels are associated with higher levels of over-reporting, as the proportion hovers between 83-85%, apart from category 3 over-reporting ~3 percentage points less than the mean of the other categories.

Additionally, the evidence that variables *Passive User* and *User Engagement Category* provide for support of H2 approaches zero, showing that active and passive users over-report at almost identical rates. When specified by *Passive User*, the sample of passive users is nearly cut in half when compared to the *User Engagement Category* specification. This shows that the findings are robust as the definition of passive users varies.

Table 2

Proportion of Over-reporting by Independent Variables of Interest

	All nonvoters	Over-reporting nonvoters	Over-reporting nonvoters (%)
Is Politically Engaged			
Yes	2941	2478	84.25
No	1243	1068	85.92
SME Score			
0	1243	1068	85.92
1	1031	872	84.58
2	763	657	86.12
3	532	433	81.39
4	325	293	83.24
5	263	223	84.79
Passive User			
Yes	1080	918	84.66
No	3104	2628	85.00
User Engagement Cat.			
Passive	2090	1788	85.55
Active	2094	1758	83.95
Ideology			
Very Liberal	615	486	79.02
Liberal	750	583	77.73
Moderate	1210	1010	83.47
Conservative	897	829	92.48
Very Conservative	572	529	92.48
Not Sure	140	109	77.85

Note: SME = Social Media Engagement

Finally, *Ideology* reveals promising findings. The proportion of over-reports moves in a seemingly linear fashion from Very Liberal with the lowest proportion of over-reporters below the mean, to Very Conservative with the highest proportion and highest above the mean of all variables.

Little evidence to support H1 or H2 are revealed in the exploratory analysis, yet some indication of support for H3 is found in the variation of over-report behavior between ideologies. However, further testing is required to unveil relationships between social media and over-reporting that may be present when controlling for other explanatory variables. To analyze these relationships further, the logistic regression models are implemented to further test for associations between political social media usage and over-reporting.

Table 2 gives average marginal effects for the independent variables of interest used in all models, showing the categorical and continuous variant of each model in tandem. The full regression output with control variables included can be found in Table A.2 of the Appendix. Model 1 sought to investigate the relationship between being politically engaged on social media and the likelihood of over-reporting vote intentions, with the regression output showing significant evidence in reverse of what was hypothesized in H1. In the model where all controls are inputted as categorical, when compared to not being politically engaged, those who indicated engagement in at least one of the social media activities are 3 percentage points less likely to over-report their intention to vote, holding all variables constant and with significance at the 1% level. The variant with continuous control variables yields similar results with a 2.7 percentage point decrease in likelihood of over-reporting, with significance at the 5% level, ceteris paribus. This study theorized that those more susceptible to the social desirability bias manifested in social media would be more likely to over-report than their inactive counterparts, but this newfound empirical evidence shows the opposite to be true. To further investigate the degree to which social media engagement affects over-reporting,

Model 2 examines how the relationship between social media and over-reporting varies across different levels of activity among those who are engaged. With neither continuous nor categorical variants showing signs of significance, the relationships between the varying levels of engagement and over-reporting behavior remain veiled, yielding little evidence in support of H2. Models 3 and 4 are implemented to provide layers of robustness for the findings of Model 2, but again show a lack of significant results.

However, the continuous variant of Model 5 yields confirmatory results that social media engagement does have some impact on one's propensity to over-report when ideology is included in the model. Specifically, for each unit increase in social media engagement, the likelihood of over-reporting decreases by 0.9 percentage points, significant at the 5% level and while holding other variables constant. Despite each level of social media engagement being insignificant in the categorical model, the inclusion of ideology in the continuous variant proves to unveil a significant linearized relationship between social media engagement and over-reporting. While effect sizes are minimal, support for H2 is nonetheless found that more passive users are more likely to over-report.

Table 2

Logistic regression models of over-reporting comparing categorical and continuous variants

Average Marginal Effects

Dep. Variable: Over Reported Vote	*Model 1*		*Model 2*		*Model 3*		*Model 4*		*Model 5*	
	Cat.	*Cts.*	*Cat.*	*Cts.*	*Cat.*	*Cts.*	*Cat.*	*Cts.*	*Cat.*	*Cts.*
Is Politically Engaged	-0.03***	-0.027**								
	(0.013)	(0.013)								
Social Media Engagement				-0.008						-0.009**
				(0.005)						(0.004)
SME 1			-						-	
SME 2			0.015						0.016	
			(0.017)						(0.017)	
SME 3			-0.033						-0.025	
			(0.02)						(0.012)	
SME 4			-0.021						-0.013	
			(0.024)						(0.023)	
SME 5			-0.016						-0.014	
			(0.026)						(0.027)	
Passive User										
Yes					0.014	0.015				
					(0.014)	(0.014)				
No					-	-				
User Engagement Cat.										
Passive							0.013	0.013		
							(0.015)	(0.014)		
Not Passive							-	-		
Ideology										0.045**
										(0.006)
Very Liberal									-	
Liberal									-0.02	
									(0.026)	
Moderate									0.049**	
									(0.024)	
Conservative									0.138***	
									(0.023)	
Very Conservative									0.123***	
									(0.026)	
Not Sure									0.033	
									(0.05)	
Constant	0.087	0.883***	0.438	1.134***	0.438	0.998***	0.45	1.015***	0.357	0.148
	(0.42)	(0.258)	(0.616)	(0.328)	(0.616)	(0.333)	(0.616)	(0.332)	(0.675)	(0.430)
Observations	4,167	4,042	2,985	2,875	2,985	2,875	2,925	2,875	2,925	2,875

Note: Standard errors are presented in parentheses. Reference categories are indicated by hyphens. SME = Social Media Engagement. Number of observations slightly fluctuates between categorical and continuous models due to the omission of certain specifications with NA values. Beta coefficients are provided for interaction terms in lieu of average marginal effects. See Model 5 discussion and table A.3 in Appendix. *p < 0.1; **p < 0.05; ***p < 0.01

Model 5 also serves to test the final hypothesis that the effects of social media engagement on over-reporting tendencies vary across the ideological spectrum. The estimate for the beta coefficient of the interaction in the continuous model is 0.021, with a lack of statistical significance while holding other variables constant. The categorical model also does not show evidence of the presence of any relationship, with none of the beta coefficients being significant (beta coefficients for the interaction estimates of the categorical model can be found in the Appendix at Table A.3). Though Model 5 lacks evidence in direct support of H3, it reveals relationships between *Ideology* and over-reporting exclusive of *Ideology*'s interaction with social media engagement. A positive and significant relationship between leaning towards conservatism and over-reporting is revealed, as a one unit increase in *Ideology* is associated with a 4.5 percentage point increase in the likelihood of over-reporting at the 5% significance level, holding other variables constant. The categorical model strengthens these findings; compared to being "Very Liberal," being "Conservative" or "Very Conservative" is associated with an increase in the likelihood of over-reporting by 13.8 and 12.3 percentage points respectively, both significant at the 1% level while holding other variables constant.

Consistent with prior research, factors like *Income, Church Attendance, Age,* and *Education* consistently influence one's likelihood to over-report. Echoing findings from Ansolabehere and Hersh (2012), these variables are associated with a greater chance of over-reporting. However, the relationships vary across model specifications. For instance, *Income* consistently influences over-reporting across all linear models, but its effect is small and lacks significance in when looking at specific categories. Church Attendance's positive link with over-reporting becomes clearer when examining specific categories that underpin this connection. The estimations for *Age* reveal that individuals older than 24 have a higher propensity to over-report, with the age group 45-54 being the most significant. Political Interest further reaffirms prior studies, showing a positive link with over-reporting. Yet, its significance is confined to models with a complete nonvoter sample. This variance could arise from diverse media preferences among nonvoters. Those considering themselves well-informed might gravitate towards traditional media rather than social media for political engagement.

As for other control variables, their relationships differ from previous findings. For instance, more years of education correspond to reduced over-reporting likelihood in this study. Contrarily, earlier studies indicated that over-reporting increases with education. Here, each additional educational tier reduces the likelihood of over-reporting by 2-3 percentage points in each model, with all being significant at the 1% level (excluding Model 5) holding other factors constant. Moreover, there's a discrepancy regarding *Age's* effect on over-reporting. While Ansolabehere and Hersh (2012) observed a negative relationship for most age groups barring 55+, this research indicates a consistent positive association, supporting the notion that older individuals are more likely to over-report.

One primary reason these findings could diverge in both direction and significance from prior research is the differing temporal perspectives from which over-reporting is analyzed. While our study takes a pre-hoc perspective, focusing on individuals who express an intent to vote but eventually do not, much of the existing literature evaluates over-reporting post-hoc, examining those who claim to have voted after an election despite not having done so.

To further explore the distinctions between these two perspectives, replicating the post-hoc dependent variable from earlier models using the new 2020 dataset would shed light on the robustness of past analyses over the subsequent 12 years. However, an analysis of the post-hoc over-reporting metric in the data revealed a significant challenge. The data shows that a staggering 99.5% of respondents validated to have not voted claimed post-election that they did. Such extreme skewness in the dependent variable renders this metric unsuitable for analysis, as using this over-reporting measure as a dependent variable would not provide robust results given the lack of variance. This stark discrepancy underscores the potential behavioral differences between pre-election intentions and post-election reporting, further highlighting the distinctiveness of this study's pre-hoc approach.

Limitations

While this research provides valuable insights into the relationship between political activity on social media and over-reporting, it is essential to recognize certain limitations that must be considered when evaluating the robustness of the methodologies used in procuring this study's results.

The documentation for the CES 2020 dataset notes that it can be misleading to draw conclusions from analyses that "cherry pick" small subsets of the data. When looking purely at small subsamples of only a few hundred observations out of the 61,000, it is possible for measurement error to lead the analysis astray. As this thesis largely relies on these subsets, other biases may be at play that could have pivotal effects on the results. To address this challenge in further research, studies should look to expand the sample size where feasible. This could mean conducting more general analyses with less specific groups, or aggregating results across multiple years in which the study was conducted to create a more representative sample. Furthermore, further abundance of CES style datasets utilizing voter validation methods would allow for cross-validation of results across datasets, improving the robustness of any inferences drawn.

Additionally, the literature on over-reporting used to support theory was based on post-hoc analysis, so this thesis assumes the explanatory demographics of over-reporting is consistent between post-hoc and pre-hoc perspectives. While the efficacy of a implementing a robustness test was discussed earlier in the results section, there is still theoretical merit in believing the two groups would behave alike. This assumption of similar over-reporting behavior holds on the basis that the crux of social desirability bias is the act of actively lying about your voting behavior. In the case of post-hoc over-reporting, the individual is often lying about whether they voted unless they somehow forgot (see Belli et al. 1999). When the individual is asked before voting day and the question is about mere intentions, there are two ways to establish lying behavior: The first is if they say that they definitely plan to vote, or if they say they have already voted early. An exception to the first case is if the individual does not vote by some sort of random error, i.e. illness, work/family emergency, natural disaster, etc. This thesis assumes that observations featuring this sort of error is the outlier due to the unlikely nature of these events combined with the explicit certainty that these individuals feel in their stated intention to vote. Since this study only analyzing individuals that said “yes definitely” or “yes, plan to vote before November 3rd” , or said they already voted, so they are more likely to be lying about it.

Conclusion

This thesis sought to better understand the phenomenon of over-reporting by taking a novel, pre-hoc perspective on the issue, and by expanding on previous empirical findings that found certain demographic traits related to social desirability bias to be primary indicators of over-reporting behavior. The 2020 US election provided an opportune angle to approach the complexity of previously unstudied demographics, not only because of the availability and comprehensiveness of the CES dataset, but also due to the prominence of political activity on social media in tandem with peak levels of ideological divide among voters. To answer the question of how political social media usage affects over-reporting behavior, subsamples of voters were analyzed to better understand how varying levels of social media, as well as social media usage across the political spectrum could affect over-reporting. It was found that individuals who are engaged in at least one type of political activity on social media have decreased chances of over-reporting. While this finding refutes the first hypothesis that those engaged in politics online are more likely to over-report than their un-engaged counterparts, the magnitude is relatively small, and possibly due to error via small sample size. Additionally, it was found with high significance that Conservatives and Very Conservatives are more likely to over-report than those considered Very Liberal with large effect, yet no effect was found when interacting *Ideology* with *Social Media Engagement*. The small effect sizes found to support the hypotheses together with the set of limitations described in the previous section suggest the need for further investigations, possibly with larger and more diverse samples to uncover more robust underlying mechanisms at play. Nonetheless, this study takes a step towards unraveling the increasingly complex relationship between digital political engagement and over-reporting. The findings shed light on the potential for social media engagement and ideology to influence over-reporting behavior and highlight the need for further research one over-reporting behavior among nonvoters.

Appendix

Table A.1

Descriptive Statistics of Control Variables			
	Frequency	Frequency (%)	Cum. Frequency (%)
Education			
No HS	72	1.72	1.72
HS Graduate	959	22.92	24.64
Some College	1351	32.29	56.93
College Graduate	1067	25.5	82.43
Postgraduate	735	17.56	100
Income			
$0-29,999	726	17.35	17.35
$30,000-49,999	784	18.74	36.1
$50,000-99,999	1357	32.43	68.52
$100,000+	924	22.08	90.61
Prefer not to say	378	9.03	100
Black			
Yes	355	8.48	8.48
No	3829	91.52	100
Other non-White			
Yes	356	8.51	8.51
No	3828	91.49	100
Married			
Yes	2291	54.75	54.75
No	1893	45.25	100
Church Attendance			
Never	1290	30.83	30.83
Few times a year	1538	36.76	67.6
Few times a month	270	6.45	74.04
At least once a week	1038	24.81	98.85
Don’t know	48	1.15	100
Age			
18-24	173	4.13	4.13
25-34	575	13.74	17.88
35-44	743	17.76	35.64
45-54	693	16.56	52.2
55+	2000	47.8	100
Ideological Strength			
Weak	1350	32.27	32.37
Moderate	1647	39.36	71.63
Strong	1187	28.37	100
Female			
Yes	2481	59.3	59.3
No	1703	40.7	100

Political Interest			
Hardly interested	188	4.49	4.49
Only now and then	324	7.74	12.24
Some of the time	1032	24.67	39.9
Most of the time	2638	63.05	100

Table A.2

Logistic regression models of over-reporting comparing categorical and continuous variants

Average Marginal Effects

Dep. Variable: Over Reported Vote	*Model 1*		*Model 2*		*Model 3*		*Model 4*		*Model 5*	
	Cat.	*Cts.*	*Cat.*	*Cts.*	*Cat.*	*Cts.*	*Cat.*	*Cts.*	*Cat.*	*Cts.*
Is Politically Engaged	-0.03***	-0.027**								
	(0.013)	(0.013)								
Social Media Engagement		-0.008								-0.009**
		(0.005)								(0.004)
SME 1			-						-	
SME 2			0.015						0.016	
			(0.017)						(0.017)	
SME 3			-0.033						-0.025	
			(0.02)						(0.012)	
SME 4			-0.021						-0.013	
			(0.024)						(0.023)	
SME 5			-0.016						-0.014	
			(0.026)						(0.027)	
Passive User					0.014	0.015				
					(0.014)	(0.014)				
User Engagement Cat.							0.013	0.013		
							(0.015)	(0.014)		
Ideology										0.045**
										(0.006)
Very Liberal									-	
Liberal									-0.02	
									(0.026)	
Moderate									0.049**	
									(0.024)	
Conservative									0.138***	
									(0.023)	
Very Conservative									0.123***	
									(0.026)	
Not Sure									0.033	
									(0.05)	
Education		-0.02***		-0.03***		-0.03***		-0.025***		-0.01*
		(0.006)		(0.006)		(0.007)		(0.006)		(0.006)
No HS	-		-		-		-		-	
HS Graduate	0.016		0.006		0.001		0.0017		-0.007	
	(0.04)		(0.05)		(0.05)		(0.052)		(0.054)	
Some College	0.007		-0.004		-0.008		-0.008		-0.001	
	(0.04)		(0.05)		(0.051)		(0.051)		(0.053)	
College Graduate	-0.012		-0.036		-0.04		-0.038		-0.028	
	(0.04)		(0.05)		(0.052)		(0.052)		(0.054)	
Postgraduate	-0.05		-0.07		-0.07		-0.074		-0.048	
	(0.04)		(0.055)		(0.053)		(0.053)		(0.055)	
Income		0.001**		0.0005*		0.0005*		0.0005*		0.0004*
		(0.0002)		(0.0003)		(0.0003)		(0.0003)		(0.0002)
$0-29,999	-		-		-		-		-	
$30,000-49,999	-0.01		-0.007		-0.009		-0.007		-0.019	
	(0.02)		(0.024)		(0.024)		(0.024)		(0.024)	
$50,000-99,999	0.0182		0.0185		0.018		0.019		0.01	
	(0.018)		(0.023)		(0.023)		(0.023)		(0.022)	
$100,000+	0.012		0.003		0.002		0.003		-0.004	
	(0.021)		(0.026)		(0.026)		(0.026)		(0.025)	
Prefer not to say	0.045**		0.043		0.041		0.044		0.031	

	(0.022)		(0.027)		(0.027)		(0.027)		(0.027)	
Black		-0.035		-0.08***		-0.08***		-0.08***		-0.011
		(0.022)		(0.03)		(0.03)		(0.03)		(0.02)
Yes	-0.015		-0.062**		-0.071**		-0.06**		-0.035	
	(0.02)		(0.028)		(0.029)		(0.029)		(0.026)	
No	-		-		-		-		-	
Other non-White		-0.003		0.0074		0.007		0.007		0.0009
		(0.02)		(0.02)		(0.02)		(0.02)		(0.019)
Yes	0.0059		0.0095		0.005		0.008		0.004	
	(0.019)		(0.023)		(0.023)		(0.023)		(0.02)	
No	-		-		-		-		-	
Married		0.024**		0.017		0.017		0.017		0.017
		(0.012)		(0.015)		(0.015)		(0.015)		(0.012)
Yes	0.03**		0.022		0.0157		0.023		0.011	
	(0.013)		(0.015)		(0.016)		(0.016)		(0.016)	
No	-		-		-		-		-	
Church Attendance		0.021***		0.02***		0.019***		0.02***		0.0057
		(0.005)		(0.006)		(0.006)		(0.006)		(0.0056)
Never	-		-		-		-		-	
Few times a year	0.055***		0.038**		0.037**		0.038**		0.012	
	(0.014)		(0.017)		(0.029)		(0.017)		(0.016)	
Few times a month	0.03		0.027		0.027		0.027		-0.019	
	(0.025)		(0.029)		(0.029)		(0.029)		(0.031)	
At least every week	0.076***		0.063***		0.063***		0.063***		0.008	
	(0.015)		(0.018)		(0.018)		0.018)		(0.021)	
Don't know	0.027		0.011		0.008		0.007		-0.025	
	(0.05)		(0.069)		(0.069)		(0.07)		(0.073)	
Age										
18-24	-	-	-		-		-		-	
25-34	0.068**	0.057	0.103**	0.09**	0.099**	0.09**	0.098**	0.089**	0.074*	0.042
	(0.035)	(0.035)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.038)	(0.032)
35-44	0.103***	0.104***	0.135***	0.132***	0.13***	0.131**	0.13***	0.13***	0.102***	0.084***
	(0.035)	(0.035)	(0.043)	(0.043)	(0.042)	(0.043)	(0.042)	(0.043)	(0.038)	0.031)
45-54	0.101***	0.113***	0.153***	0.154***	0.149***	0.153***	0.149***	0.152***	0.108***	0.082***
	(0.035)	(0.035)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.39)	(0.032)
55+	0.06*	0.059*	0.087***	0.085**	0.083***	0.089**	0.083*	0.083**	0.037	0.026
	(0.035)	(0.034)	(0.043)	(0.043)	0.042)	(0.043)	(0.042)	(0.04)	(0.038)	(0.03)
Ideological Strength		0.01		0.0044		0.0038		0.004		0.024***
		(0.007)		(0.007)		(0.0089)		(0.009)		(0.008)
Weak	-		-		-		-			
Moderate	0.026**		0.014		0.013		0.013			
	(0.0135)		(0.017)		(0.017)		(0.017)			
Strong	0.025*		0.0122		0.011		0.011			
	(0.015)		(0.018)		(0.018)		(0.018)			
Female		0.007		0.0032		0.0033		0.003		0.014
		(0.012)		(0.014)		(0.0137)		(0.014)		(0.018)
Yes	0.009		0.007		0.005		0.005		0.019	
	(0.012)		(0.014)		(0.014)		(0.014)		(0.014)	
No	-		-		-		-		-	
Political Interest		0.022***		0.012		0.001		0.001		0.027***
		(0.008)		(0.012)		(0.01)		(0.01)		(0.008)
Hardly interested	-		-		-		-		-	
Only now and then	0.073*		0.015		0.014		0.015		0.022	
	(0.04)		(0.068)		(0.067)		(0.067)		(0.069)	
Some of the time	0.12***		0.05		0.049		0.05		0.05	
	(0.037)		(0.06)		(0.061)		(0.061)		(0.06)	
Most of the time	0.133***		0.069		0.065		0.065		0.08	
	(0.038)		(0.061)		(0.06)		(0.06)		(0.06)	
Constant	0.087	0.883***	0.438	1.134***	0.438	0.998***	0.45	1.015***	0.357	0.148
	(0.42)	(0.258)	(0.616)	(0.328)	(0.616)	(0.333)	(0.616)	(0.332)	(0.675)	(0.430)
Observations	4,167	4,042	2,985	2,875	2,985	2,875	2,925	2,875	2,925	2,875

Note: Standard errors are presented in parentheses. Reference categories are indicated by hyphens. *Ideological strength* is dropped from Model 5 (Cat.) due to collinearity concerns. p < 0.1; **p < 0.05; ***p < 0.01

Table A.3

Logistic regression model output of interaction effects for Model 5 (Cat.)			
β *coefficients*			
Dep. Variable: Over-reported Vote	*Model 5 (Cat.)*		
SME 1 * Liberal	-	SME 1 * Very Conservative	-
SME 2 * Liberal	-0.036 (0.411)	SME 2 * Very Conservative	-0.627 (0.584)
SME 3 * Liberal	-0.558 (0.420)	SME 3 * Very Conservative	-0.629 (0.624)
SME 4 * Liberal	-0.226 (0.480)	SME 4 * Very Conservative	-0.029 (0.671)
SME 5 * Liberal	0.156 (0.536)	SME 5 * Very Conservative	0.561 (0.789)
SME 1 * Moderate	-	SME 1 * Not sure	-
SME 2 * Moderate	-0.213 (0.402)	SME 2 * Not sure	0.556 (0.935)
SME 3 * Moderate	-0.666 (0.425)	SME 3 * Not sure	0.319 (0.955)
SME 4 * Moderate	-0.043 (0.473)	SME 4 * Not sure	12.764 (264.14)
SME 5 * Moderate	0.332 (0.608)	SME 5 * Not sure	0.603 (1.3)
SME 1 * Conservative	-		
SME 2 * Conservative	-0.001 (0.502)		
SME 3 * Conservative	0.031 (0.565		
SME 4 * Conservative	0.390 (0.661)		
SME 5 * Conservative	0.346 (0.354)	Constant	0.357 (0.675)
		Observations	2,925

Note: SME = Social Media Engagement

Bibliography

Ansolabehere, Stephen, and Hersh, Eitan. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." *Political Analysis*, vol. 20, no. 4, 2012, pp. 437–459, https://doi.org/10.1093/pan/mps023.

Bail, Christopher A. "Political Sectarianism in America | Science." *Science*, 30 Oct. 2020, www.science.org/doi/10.1126/science.abe1715.

Barberá, Pablo. "Social Media, Echo Chambers, and Political Polarization." *Social Media and Democracy*, 2020, pp. 34–55, https://doi.org/10.1017/9781108890960.004.

Belli, Robert F., et al. "Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly*, vol. 63, no. 1, 1999, p. 90, https://doi.org/10.1086/297704.

Bernstein, Robert, et al. "Overreporting Voting." *Public Opinion Quarterly*, vol. 65, no. 1, 2001, pp. 22–44, https://doi.org/10.1086/320036.

Brenner, Philip S., and John DeLamater. "Lies, Damned Lies, and Survey Self-Reports? Identity as a Cause of Measurement Bias." *Social Psychology Quarterly*, vol. 79, no. 4, 2016, pp. 333–354, https://doi.org/10.1177/0190272516628298.

Cantril, Hadley, and John Harding. "The 1942 Elections: A Case Study in Political Psychology." *Public Opinion Quarterly*, vol. 7, no. 2, 1943, p. 222, https://doi.org/10.1086/265616.

Clausen, Aage R. "Response Validity: Vote Report." *Public Opinion Quarterly*, vol. 32, no. 4, 1968, p. 588, https://doi.org/10.1086/267648.

Dixon, Stacy Jo. "U.S. Social Platform User Engagement Growth 2020." *Statista*, 30 Mar. 2022, www.statista.com/statistics/265069/social-platform-user-engagement-change-usa/.

Drakulich, Kevin, and Megan Denver. "The Partisans and the Persuadables: Public Views of Black Lives Matter and the 2020 Protests." Perspectives on Politics 20.4 (2022): 1191-208. Print. https://doi.org/10.1017/S1537592721004114

Duff, B., et al. "Good Excuses: Understanding Who Votes with an Improved Turnout Question." *Public Opinion Quarterly*, vol. 71, no. 1, 2007, pp. 67–90, https://doi.org/10.1093/poq/nfl045.

Jackman, Mary R., and Michael J. Muha. "Education and Intergroup Attitudes: Moral Enlightenment, Superficial Democratic Commitment, or Ideological Refinement?"

American Sociological Review, vol. 49, no. 6, 1984, p. 751, https://doi.org/10.2307/2095528.

Kuha, Jouni. "Why Are Polls Important during Elections?: LSE Research." *Why Are Polls Important during Elections? | LSE Research*, www.lse.ac.uk/research/research-for-the-world/impact/the-politics-of-polling-why-are-polls-important-during-elections.

Grimm, Pamela. "Social Desirability Bias." *Wiley International Encyclopedia of Marketing*, 2010, https://doi.org/10.1002/9781444316568.wiem02057.

Hanmer, Michael J., et al. "Experiments to Reduce the Over-Reporting of Voting: A Pipeline to the Truth." *Political Analysis*, vol. 22, no. 1, 2014, pp. 130–141, https://doi.org/10.1093/pan/mpt027.

Kennedy, Courtney. "Confronting 2016 and 2020 Polling Limitations." *Pew Research Center Methods*, 8 Apr. 2021, www.pewresearch.org/methods/2021/04/08/confronting-2016-and-2020-polling-limitations/.

Karp, Jeffrey A., and David Brockington. "Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries." *The Journal of Politics*, vol. 67, no. 3, 2005, pp. 825–840, https://doi.org/10.1111/j.1468-2508.2005.00341.x.

Krämer, Nicole C., and Stephan Winter. "Impression Management 2.0." *Journal of Media Psychology*, vol. 20, no. 3, 2008, pp. 106–116, https://doi.org/10.1027/1864-1105.20.3.106.

Miller, Jon D et al. "The impact of the Covid-19 pandemic in a polarized political system: Lessons from the 2020 election." *Electoral studies* vol. 80 (2022): 102548. doi:10.1016/j.electstud.2022.102548

Morin-Chassé, Alexandre, et al. "How to Survey about Electoral Turnout? The Efficacy of the Face-Saving Response Items in 19 Different Contexts." *Political Science Research and Methods*, vol. 5, no. 3, 2016, pp. 575–584, https://doi.org/10.1017/psrm.2016.31.

Piatak, Jaclyn, and Ian Mikkelsen. "Does Social Media Engagement Translate to Civic Engagement Offline?" *Nonprofit and Voluntary Sector Quarterly*, vol. 50, no. 5, 2021, pp. 1079–1101, https://doi.org/10.1177/0899764021999444.

Silver, Brian D., et al. "Who Overreports Voting?" *American Political Science Review*, vol. 80, no. 2, 1986, pp. 613–624, https://doi.org/10.2307/1958277.

Suciu, Peter. "Social Media Could Determine the Outcome of the 2020 Election." *Forbes*, 10 Dec. 2021, www.forbes.com/sites/petersuciu/2020/10/26/social-media-could-determine-the-outcome-of-the-2020-election/?sh=76765f4926f6