Biases of Online Political Polls: Who Participates?

Eszter Hargittai1 and Gokce Karaoglu1

Abstract
With a large portion of the population online and the high cost of phone-based surveys, querying people about their voter preference online can offer an affordable and timely alternative. However, given that there are biases in who adopts various sites and services that are often used as sampling frames (e.g., various social media), online political polls may not represent the views of the overall population. How are such polls biased? Who is most likely to participate in them? Drawing on a national survey of voter-eligible American adults administered in summer 2016, this paper shows that background characteristics (i.e., age, gender, race, education, and employment status) as well as Internet experiences and skills relate to who casts votes in online political polls.

Keywords
political polls, online political polling, polling biases, survey, sampling bias, Internet skills

Introduction
As Internet use has spread to an increasing portion of the population (Anderson, Perrin, and Jiang 2018), online polling has garnered more popularity (Mortimore and Wells 2017). This is understandable given that online surveys can be cost effective, especially in terms of administration since they do not require interview personnel (in the case of phone-based or in-person data collection) or postage logistics (Evans and Mathur 2005). Conducting a survey online may make it available to a greater number of people than an in-person survey and speeds up the distribution and collection process (Couper 2000). However, online surveys also have weaknesses, not least of which is that not everybody uses the Internet; moreover, even among those who do, people vary considerably in their online behavior. In this paper, we analyze survey data about a diverse group of voter-eligible American adults to examine what explains who casts votes in online political polls, with the goal of shining light on what types of biases such polls may include.

Challenges of Online Political Polling
Reaching a representative sample of the voter-eligible population is a challenge for political polls, both online and off (Cook, Heath, and Thompson 2000; Diaz et al. 2016; Huberty 2015; Kellner, Twyman, and Wells 2011). The classic example is that of the Literary Digest Poll concerning the 1936 U.S. presidential election. The paper was confident in its predictions due to the sheer number of polls it had collected through postal mail, reminiscent of today’s excitement about big data (e.g., Anderson 2008). However, basing its sampling on its own above-mean-income subscriber list as well as telephone and car ownership, the Literary Digest Poll systematically excluded less privileged populations from participation, namely, those more likely to vote for the candidate who ultimately won despite the poll’s predictions (Squire 1988).

Although Internet use is widespread in the United States, 13 percent of adults were not online in 2016, a group mostly consisting of seniors, the less educated, and those who make less than $30,000 annually (Anderson and Perrin 2016). Even if connected, people engage in very different activities online, and what they do is related to several sociodemographic factors (Robinson et al. 2015). For example, people from less privileged backgrounds are more likely to use the Internet for entertainment types of activities compared to higher rates of capital-enhancing usage by people of higher socioeconomic status (Bonfadelli 2002; Eynon 2009; Hale et al. 2010; Zillien and Hargittai 2009). Topical interest has
also been tied to the adoption of particular online services such as the role of being interested in celebrity news and entertainment in driving Twitter adoption (Hargittai and Litt 2011). Such research suggests that even among those online, people may not be equally inclined to participate in a poll. This is the question this paper addresses.

When considering online polling about political topics in particular, an important question to ask is whether Internet users are representative of the general population regarding their political interests and voting preferences. Research has shown this not to be the case (Zhang et al. 2009). In the 2008 U.S. presidential election, online platforms were extensively used by Obama campaigners, but Republicans, and McCain supporters in particular, were far less active on Twitter (Gayo-Avello 2011). Online political participation relates to some extent to liberal attitudes and higher socioeconomic status (Best and Krueger 2005), suggesting a biased sample when it comes to whose voices may be represented in such polls.

A challenge of political polling more generally is social desirability bias. People may not respond to questions accurately if they believe that their answer is not socially desirable (Crowne and Marlowe 1960; Silver, Anderson, and Abramson 1986). During the 2016 U.S. presidential election campaign, the news media increasingly covered Trump’s candidacy in a negative light, often making derogatory remarks (Patterson 2016). Given such widespread adverse coverage, it is possible that some Trump supporters would be shy about expressing their support for his presidency. A study of college students’ voting preference conducted right before the 2016 U.S. presidential election found that those more prone to controlling their public appearances to impress others were less likely to say they support Trump when controlling for demographic and socioeconomic factors (Klar, Weber, and Krupnikov 2016). Another study looking at national as well as state-level data found evidence of “hidden Trump supporters” (Enns, Lagodny, and Schuldt 2017), namely, people who supported Trump but not necessarily expressed their opinions in a survey when asked about their vote intentions. Yet, conducting an extensive analysis of the performance of pre-election polls in 2016, a study found little evidence of social desirability bias, therefore suggesting the effect on the poll error to be relatively small (Kennedy et al. 2018).

Addressing the challenge of social desirability bias, some research on mode effects has suggested that people are less prone to social desirability effects in online surveys than in-person or telephone interviews (Holbrook and Krosnick 2010; Kreuter, Presser, and Tourangeau 2008). The absence of an interviewer becomes an important advantage of online surveys, especially when the context includes sensitive topics (Duffy et al. 2005). Survey mode, for that reason, might have an effect on people’s desirability to give honest answers when the questions concern opinions about a controversial figure such as Donald Trump. From this perspective, online political polls may offer a better method than other modes for asking people about their voter preference if they are less prone to social desirability bias. Kennedy and colleagues (2017) did not find this to be the case, however, when looking at public support for key policy proposals during Trump’s presidency.

In sum, online polls look like a promising alternative to traditional types of surveys as long as the sample is representative of the target population. Given that reaching a representative sample is still an issue due to various reasons discussed earlier concerning sociodemographic biases, understanding in which ways online poll participants skew is an important step toward solving the potential problems that come with it.

Data and Methods

To examine what biases may go into who takes online political polls, we turn to a survey that had very little content about politics and mostly focused on other issues such as the types of social media people use and whether they contribute to user-generated sites like Wikipedia. This survey was administered online, and the irony of using this mode of data collection to explore mode-based biases in political polls is not lost on the authors. Regarding the biases we explore, however, our findings are likely to be conservative given our mode of data collection. We recognize the limitations our approach poses and discuss them in detail later in the paper.

Data Collection

We draw on unique survey data of a national sample of U.S. adults 18 years old and over, collected in summer 2016, a few months before the 2016 presidential election. We contracted with the independent research organization National Opinion Research Center at the University of Chicago (NORC subsequently) to administer questions to their AmeriSpeak panel online. AmeriSpeak is a nationally representative, probability-based survey panel (National Opinion Research Center 2017). After pretesting the survey with 23 respondents and updating items based on the results in early May 2016, we ran the survey May 25 through July 5, 2016. Of note is that by the time of the survey, Hillary Clinton and Donald Trump were the presumptive nominees of their respective parties. For survey quality, we included an attention-check question and only analyze responses from participants who passed this question. In total, the sample includes valid responses from 1,512 American adults 18 and over, which constitutes a 37.8 percent survey response rate. The analyses we present here concern the 1,441 voting-eligible respondents.

Measures: Independent Variables

Demographic and socioeconomic factors. Background variables about respondents such as their age, gender, education,
income, and race/ethnicity were supplied by NORC based on their earlier data collection about the AmeriSpeak panel. Here we describe what coding we used for these variables. We report age as a continuous variable. We created three education categories: high school or less, some college, and college degree or more. Income was reported in 18 categories, which we recoded to their midpoint values to make it a continuous variable. In the regression analyses, we use the log of income. Race and ethnicity are dummy variables for white, Hispanic, African American, Asian American, Native American, and other. We created a dichotomous “coupled” measure for those either married or living with a partner. We have a dummy variable for those employed either full-time or part-time. We also have a dummy variable signaling rural residence. Finally, there is a continuous measure of household size.

**General Internet experiences and skills.** We include measures for how long people have been Internet users, how much autonomy they have in freely accessing the Internet when and where they want to, how much time they spend online, and their Internet skills, variables that prior literature has found important in understanding people’s online experiences (DiMaggio et al. 2004). We also control for Internet use for political purposes.

We asked respondents when they had first started using the Internet, offering the following answer options with their recoded values in parentheses: within the past year (1), 1 to 5 years ago (2.5), more than 5 but less than 10 years ago (7.5), and 10 or more years ago (12.5). To measure autonomy of use, we asked, “At which of these locations do you have access to the Internet, that is, if you wanted to you could use the Internet at which of these locations?” followed by nine options such as home, workplace, and friend’s home. To assess frequency of use, we asked, “On an average weekday, not counting time spent on email, chat and phone calls, about how many hours do you spend visiting Web sites?” and then asked the same question about “average Saturday or Sunday.” The answer options ranged from none to 6 hours or more (9.7 percent) of respondents were thus excluded. Assuming the analyses presented in the following with the excluded people (these are questions not asked on this specific survey but supplied by AmeriSpeak from their data about the panel), those excluded are more likely to be lower educated (p < .01), have a lower income (p < .05), and be of Hispanic origin (p < .001) or African American (p < .001). Those included are more likely to be white (p < .001). There is no difference by gender, being Asian American or Native American, being coupled, or living in a rural area. Running the analyses presented in the following with the excluded cases yields similar results regarding significance of the various variables, with some variation in size of the coefficients.

**Political ideology.** The survey asked respondents how they would describe their political views with the following answer options: very liberal, fairly liberal, middle of the road, fairly conservative, very conservative, and other. These responses were recoded into three dummies: liberal (either very liberal or liberal), middle of the road, and conservative (either very conservative or conservative).

**Measures: Dependent Variable**

To measure participation in online political polling, we asked: “Have you ever done any of the following?” with numerous online actions listed, one of which was: “Submitted a vote to an online political poll.” While this no/yes binary question may seem too simplistic, the majority (59.3 percent) had never done so, making it a helpful measure of online political poll participation.

**The Sample**

Table 1 shows the sample characteristics. Almost the same number of men and women participated; the average age is 48.9. Just under 30 percent of respondents were ethnic and racial minorities (11.6 percent Hispanic, 11.5 percent African American, 2.9 percent Asian American, 1.7 percent Native American, and .8 percent other.) A quarter have no more than a high school education, just under a third (31.4 percent) completed some college, and 43.5 percent have at least a college degree. Average income at $72,091 is higher than the national average. The mean household size is 2.6, just under 63 percent are employed, 61.4 percent are coupled, and 13.4 percent live in a rural area. Regarding their online experiences, respondents on average have been using the Internet for just over 11 years, have 4.8 locations where they can go online, and spend 14.7 hours on the Web weekly. Their Internet skills are varied; on a 1 to 5 scale, they average a 3.4 score.

Including those in the sample who had not passed the attention-check question would add noise to the study since we cannot assume that their responses to other questions on the survey were not similarly error-prone. About 10 percent (9.7 percent) of respondents were thus excluded. Assuming demographic information in AmeriSpeak is correct about these people (these are questions not asked on this specific survey but supplied by AmeriSpeak from their data about the panel), those excluded are more likely to be lower educated (p < .01), have a lower income (p < .05), and be of Hispanic origin (p < .001) or African American (p < .001). Those included are more likely to be white (p < .001). There is no difference by gender, being Asian American or Native American, being coupled, or living in a rural area. Running the analyses presented in the following with the excluded cases yields similar results regarding significance of the various variables, with some variation in size of the coefficients.

**Analyses**

First, we present bivariate relationships of respondent background characteristics and having taken an online political
Results

The bivariate analyses in Table 2 show that the oldest quartile of respondents were significantly more likely to have taken an online political poll than the youngest. Men have done so significantly more than women, and whites significantly more than people of other racial and ethnic backgrounds (except the “other” category). Both education and income are positively related to participating in an online political poll. Household size matters (the smaller, the more likely to have participated), as does rural residence (less likely than others), but being employed and coupled do not make a difference to taking an online political poll.

People’s Internet experiences are very much related to participation in online political polls, with those who have been Internet users longer, who have more autonomy of use, who spend more time online, and with higher skills all more likely to have engaged in this activity. Regarding political leaning, bivariate analyses suggest that liberals are much more likely to have participated in online political polls while those in the middle of the road are much less likely than conservatives. Next, we turn to whether these results are robust when controlling for other factors.

The logistic regression results in Table 3 first show how sociodemographic factors relate to online polling participation (Model 1); next, we include Internet experiences and skills in the model (Model 2), and finally, we also look at political ideology while controlling for the aforementioned factors.
factors (Model 3). The first model shows that gender, race/ethnicity, and education are all related to the likelihood of having taken an online political poll: Women, Asian Americans, African Americans, and those with no more than a high school education were all less likely to have done so. The second model suggests that autonomy of use, amount of time spent online, and Internet skills are all positively related to online political poll participation. Once we control for these factors, older adults are also more likely to have participated.

Once we add political ideology to the model (Model 3), the various sociodemographics thus far significant remain so. We also find that being liberal versus conservative does not make a difference for having taken an online political poll. However, those who classify themselves as middle of the road are much less likely to have participated in such a poll. We discuss the implications of these findings in the next section.

### Discussion and Conclusion

A major limitation of the present work is that the survey on which it relies was itself collected online while attempting to identify biases that go into who participates in online political polls. The ideal mode of data collection would be one that is not related to the methodological issues the paper itself tackles: biases in taking online polls. That said, absent other available data sets with the relevant covariates and dependent variable, we believe the paper nonetheless makes relevant contributions to the political polling literature. We believe that despite its limitations, the study offers a unique opportunity to answer the question we raise: What types of Internet users are most likely to participate in online political polls?

Our findings suggest that certain demographic and socioeconomic characteristics such as being male and white, having higher education, as well as spending more time online, having more autonomy of use, and having higher Internet skills are all positively related to the likelihood of taking an online political poll. In terms of ideology, people identifying themselves as middle-of-the-road are less likely to participate in online political polls.

Research has noted that people with middle-of-the-road political views are affected by ideas rather than ideologies (Luntz 2008). These so-called swing voters or floaters are usually only about 50 percent certain of whom they will vote for on election day (Luntz 2008) and are more likely than...
others to show up as undecided voters in pre-election polls. As they make up their minds at the last minute, pre-election polls might not appeal to them as much, decreasing their participation in such political polls and thereby having their perspectives underrepresented in such data sets. Indeed, reviewing data from exit polls and a callback study by the Pew Research Center, Kennedy and colleagues (2018) found evidence of a late swing in vote preference toward Trump in the 2016 elections, particularly in states that Clinton lost by the smallest margins.

Recognizing the biases in who participates in online political polls is crucial if they are meant to reach a representative sample of the voter-eligible population. Our findings highlight characteristics of Internet users that correlate with likelihood to take online political polls. Future such polls administered online will want to keep these biases in mind as they analyze their data to make sure that they do not derive wrong conclusions due to selection biases.

Acknowledgments

A version of this paper was presented at the Midwest Political Science Association Annual Meetings in 2017, and the authors are grateful for the feedback received there. The authors thank Aaron Shaw and Sam Mandelsohn for their input on data collection. Hargittai is grateful to the April McClain Delaney and John Delaney Research Professorship at Northwestern University for the time made available to do this work.

ORCID iD

Eszter Hargittai https://orcid.org/0000-0003-4199-4868

References

Anderson, Chris. 2008. “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” Wired. Retrieved July 16, 2018 (https://www.wired.com/2008/06/pb-theory/).

Anderson, Monica, and Andrew Perrin. 2016. “13% of Americans Don’t Use the Internet. Who Are They?” Pew Research Center. Retrieved July 25, 2018 (http://web.archive.org/web/20161216171235/http://www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-the-internet-who-are-they/).

Anderson, Monica, Andrew Perrin, and Jingjing Jiang. 2018. “11% of Americans Don’t Use the Internet. Who Are They?” Pew Research Center. Retrieved July 2, 2018 (http://www.pewresearch.org/fact-tank/2018/03/05/some-americans-dont-use-the-internet-who-are-they/).

Best, Samuel J., and Brian S. Krueger. 2005. “Analyzing the Representativeness of Internet Political Participation.” Political Behavior 27(2):183–216.

Bonfadelli, Heinz. 2002. “The Internet and Knowledge Gaps: A Theoretical and Empirical Investigation.” European Journal of Communication 17(1):65–84.

Cook, Colleen, Fred Heath, and Russel L. Thompson. 2000. “A Meta-analysis of Response Rates in Web- or Internet-based Surveys.” Educational and Psychological Measurement 60(6):821–36.

Couper, Mick P. 2000. “Review: Web Surveys: A Review of Issues and Approaches.” The Public Opinion Quarterly 64(4):464–94.

Crowne, Douglas P., and David Marlowe. 1960. “A New Scale of Social Desirability Independent of Psychopathology.” Journal of Consulting Psychology 24(4):349.

Díaz, Fernando, Michael Gamon, Jake M. Hofman, Emre Kiciman, and David Rothschild. 2016. “Online and Social Media Data as an Imperfect Continuous Panel Survey.” PloS One 11(1):e0145406.

DiMaggio, Paul, Eszter Hargittai, Coral Celeste, and Steven Schafer. 2004. “Digital Inequality: from Unequal Access to Differentiated Use.” Pp. 355–400 in Social Inequality, edited by K. Neckerman. New York: Russell Sage Foundation.

Duffy, Bobby, Kate Smith, George Terhanian, and John Bremer. 2005. “Comparing Data from Online and Face-to-face Surveys.” International Journal of Market Research 47(6):615–39.

Emns, Peter K., Julius Lagodny, and Jonathon P. Schultd. 2017. “Understanding the 2016 US Presidential Polls: The Importance of Hidden Trump Supporters.” Statistics, Politics and Policy 8(1).

Evans, Joel R., and Anil Mathur. 2005. “The Value of Online Surveys.” Internet Research 15(2):195–219.

Eynon, R. 2009. “Mapping the Digital Divide in Britain: Implications for Learning and Education.” Learning, Media, and Technology 34(4):277–90.

Gayo-Avello, Daniel. 2011. “Don’t Turn Social Media into Another ‘Literary Digest’ Poll.” Communications of the ACM 54(10):121–28.

Hale, Timothy M., Shelia R. Cotten, Patricia Drentea, and Melinda Goldner. 2010. “Rural-urban Differences in General and Health-related Internet Use.” American Behavioral Scientist 53(9):1304–25.

Hargittai, Eszter, and Y. P. Hsieh. 2012. “Succinct Survey Measures of Web-use Skills.” Social Science Computer Review 30(1):95–107.

Hargittai, Eszter, and Eden Litt. 2011. “The Tweet Smell of Celebrity Success: Explaining Variation in Twitter Adoption among a Diverse Group of Young Adults.” New Media & Society 13(5):824–42.

Holbrook, Allyson L., and Jon A. Krosnick. 2010. “Social Desirability Bias in Voter Turnout Reports Tests Using the Item Count Technique.” Public Opinion Quarterly 74(1):37–67.

Huberty, Mark. 2015. “Can We Vote with Our Tweet? On the Perennial Difficulty of Election Forecasting with Social Media.” International Journal of Forecasting 31(3):992–1007.

Kellner, Peter, Joe Twyman, and Anthony Wells. 2011. “Polling Voting Intentions.” Pp. 94–108 in Political Communication in Britain, edited by D. Wring, R. Mortimore, and S. Atkinson. London: Palgrave Macmillan.

Kennedy, Courtney, Mark Blumenthal, Scott Clement, JoshUA d. Clinton, Claire Durand, Charles Franklin, Kyley McGeeeny, Lee Miringoff, Kristen Olson, Doug Rivers, Lydia Saad, Evans Witt, and Chris Wlezien. 2018. “An Evaluation of the 2016 Election Polls in the United States.” Public Opinion Quarterly 82(1):1–33.

Kennedy, Courtney, Scott Keeter, Andrew Mercer, Nick Hatley, Nick Bertoni and Arnold Lau. 2017. “Are Telephone Polls Understating Support for Trump?” Pew Research Center.
Retrieved July 2, 2018 (http://www.pewresearch.org/2017/03/31/are-telephone-polls-understating-support-for-trump/).

Klar, Samara, Christopher R. Weber, and Yanna Krupnikov. 2016. “Social Desirability Bias in the 2016 Presidential Election.” The Forum 14(4):433–443.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. “Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity.” Public Opinion Quarterly 72(5):847–65.

Luntz, Frank. 2008. “The Float Vote.” Los Angeles Times, June 22.

Mortimore, Roger, and Anthony Wells. 2017. “The Polls and Their Context.” Pp. 19–38 in Political Communication in Britain, edited by D. Wring, R. Mortimore, and S. Atkinson. London: Palgrave Macmillan.

National Opinion Research Center. 2017. “AmeriSpeak: NORC’s Breakthrough Panel-based Research Platform.” Retrieved July 16, 2018 (http://www.norc.org/Research/Capabilities/pages/amerispeak.aspx).

Patterson, Thomas E. 2016. News Coverage of the 2016 Presidential Primaries: Horse Race Reporting Has Consequences. Rochester, NY: Social Science Research Network.

Robinson, Laura, Shelia R. Cotten, Hiroshi Ono, Anabel Quan-Haase, Gustavo Mesch, Wenhong Chen, Jeremy Schulz, Timothy M. Hale, and Michael J. Stern. 2015. “Digital Inequalities and Why They Matter.” Information, Communication & Society 18(5):569–82.

Silver, Brian D., Barbara A. Anderson, and Paul R. Abramson. 1986. “Who Overreports Voting?” American Political Science Review 80(2):613–24.

Squire, Peverill. 1988. “Why the 1936 Literary Digest Poll Failed.” Public Opinion Quarterly 52(1):125–33.

Wasserman, Ira M., and Marie Richmond-Abbott. 2005. “Gender and the Internet: Causes of Variation in Access, Level, and Scope of Use.” Social Science Quarterly 86(1):252–70.

Zhang, Chan, Mario Callegaro, Melanie Thomas, and Charles DiSogra. 2009. “Do We Hear Different Voices?: Investigating the Differences between Internet and Non-Internet Users on Attitudes and Behaviors.” Pp. 6063–76 in Proceedings of the Section on Survey Research Methods. Alexandria, VA: American Statistical Association.

Zillien, Nicole, and Eszter Hargittai. 2009. “Digital Distinction: Status-specific Types of Internet Usage.” Social Science Quarterly 90(2):274–91.

Author Biographies

Eszter Hargittai is professor and chair of Internet Use & Society at the Institute of Communication and Media Research of the University of Zurich. Her research looks at how people may benefit from their digital media uses, with a particular focus on how differences in people’s Web-use skills influence what they do online. She is editor of the forthcoming Handbook of Digital Inequality.

Gokce Karaoglu is a PhD student and research assistant in the Internet Use & Society division at the Institute of Communication and Media Research, University of Zurich. She received her master’s degree in management and economics at the University of Zurich after receiving bachelor’s degrees in media and visual arts as well as business administration from Koç University, Istanbul. Her research interests include new media technologies, online participation, and social networks.