The Effects of Facebook and Instagram on the 2020 Election: A Deactivation Experiment

Hunt Allcott,[1,18] Matthew Gentzkow,[1,18*]
Winter Mason,[2,19,20] Arjun Wilkins,[2,19] Pablo Barberá,[2] Juan Carlos Cisneros,[1]
Adriana Crespo-Tenorio,[2] Drew Dimmery,[2] Deen Freelon,[3]
Sandra González-Bailón,[4] Andrew M. Guess,[5] Young Mie Kim,[6] David Lazer,[7]
Neil Malhotra,[8] Devra Moehler,[2] Sameer Nair-Desai,[1] Houda Nait El Barj,[1]
Brendan Nyhan,[9] Ana Carolina Paixao de Queiroz,[1] Jennifer Pan,[10] Jaime Settle,[11]
Emily Thorson,[12] Rebekah Tromble,[13] Carlos Velasco Rivera[2],
Benjamin Wittenbrink,[1] Magdalena Wojcieszak,[14,15] Saam Zahedian,[1] Annie Franco,[2,20]
Chad Kiewiet de Jonge,[2,20] Natalie Jomini Stroud,[16,21] Joshua A. Tucker[17,21]

[1]Stanford University

[2]Meta

[3]UNC Hussman School of Journalism and Media, University of North Carolina at Chapel Hill

[4]Annenberg School for Communication, University of Pennsylvania

[5]Department of Politics and School of Public and International Affairs, Princeton University

[6]School of Journalism and Mass Communication, University of Wisconsin-Madison

[7]NULab of Texts, Maps, and Networks, Northeastern University

[8]Graduate School of Business, Stanford University

[9]Department of Government, Dartmouth College

[10]Department of Communication, Stanford University

[11]Department of Government, William & Mary

[12]Department of Political Science, Syracuse University

[13]School of Media and Public Affairs and Institute for Data, Democracy, and Politics, George Washington University

[14]Department of Communication, University of California, Davis

[15]Amsterdam School of Communication Research, University of Amsterdam

[16]Moody College of Communication, University of Texas at Austin

[17]Wilf Family Department of Politics and Center for Social Media and Politics, New York University

[18]Co-lead author with final control rights

[19]Meta lead author

[20]Meta research lead

[21]Co-last author and academic research lead

*To whom correspondence should be addressed; E-mail: gentzkow@stanford.edu.

We study the effect of Facebook and Instagram access on political beliefs, attitudes, and behavior by randomizing a subset of 19,857 Facebook users and 15,585 Instagram users to deactivate their accounts for six weeks before the 2020 U.S. election. We report four key findings. First, both Facebook and Instagram deactivation reduced an index of political participation (driven mainly by reduced participation online). Second, Facebook deactivation had no significant effect on an index of knowledge, but secondary analyses suggest that it reduced knowledge of general news while possibly also decreasing belief in misinformation circulating online. Third, Facebook deactivation may have reduced self-reported net votes for Trump, though this effect does not meet our preregistered significance threshold. Finally, the effects of both Facebook and Instagram deactivation on affective and issue polarization, perceived legitimacy of the election, candidate favorability, and voter turnout were all precisely estimated and close to zero.

Introduction

Many people argue that social media has profoundly affected democracy in America. Social media may increase political polarization, in part by facilitating ideological echo chambers among like-minded people (*1–6*). It may make voters more (*5, 7–9*) or less (*10–13*) knowledgeable, depending on the balance of accurate and inaccurate information that they receive. It may either increase (*4, 14, 15*) or decrease (*16, 17*) political engagement and voter turnout, and it may impact overall trust and confidence in the electoral system (*18, 19*). Observers on each side of the political spectrum speculate that social media may tilt elections in favor of their opponents (*20–23*). Much of this discussion has focused on Facebook, but Instagram is now a more widely used platform among younger voters and contains a significant amount of political

content (*24, 25*).

These concerns were particularly relevant in the 2020 U.S. presidential election. Affective polarization among U.S. voters reached an all-time high in 2020 (*26*). Popular media articles argued that misinformation on social media was a key driver of the 2016 election outcome (*20, 21, 27*), and several observers speculated that its impact in 2020 could be as large or larger (*28–30*). Concerns about the integrity of the electoral process were widely aired on social media, including concerns about fraud and vote-by-mail (*31*), and social media played an important role in the events following the election, including the "Stop the Steal" movement questioning the election outcome and the storming of the U.S. Capitol on January 6 (*32, 33*). Some have highlighted ways that social media may have advantaged Joe Biden (*34, 35*); others have emphasized ways it benefited Donald Trump (*36*).

We use a randomized experiment to measure the effects of access to Facebook and Instagram on individual-level political outcomes during the 2020 election. We recruited 19,857 Facebook users and 15,585 Instagram users who used the platform at least 15 minutes per day at baseline. We randomly assigned 27 percent to a treatment group that was paid to deactivate their Facebook or Instagram accounts for the six weeks before election day, and the remainder to a control group that was paid to deactivate for just one week. We estimate effects of deactivation on consumption of other apps and news sources, factual knowledge, political polarization, perceived legitimacy of the election, political participation, and candidate preferences.

Existing experimental evidence on the effects of social media access is limited. Most related are earlier experiments that have measured the effect of Facebook deactivation on political outcomes in the U.S. (*5, 37*) and other countries (*38*). Ours is the first deactivation experiment to study the impact of social media outcomes in a U.S. presidential election and also the first to study the effect of Instagram. Our Facebook sample alone is more than 10 times larger than any previous deactivation experiment we are aware of. Our study also improves on ear-

lier experiments by incorporating directly-measured outcomes and administrative platform data from Meta, including a wider range of survey-based outcomes, and using a longer deactivation period. Our study builds on a broader literature using randomized experiments (*39–41*) and observational "natural experiments" (*4, 42–48*) to study political effects of social media.

Our study has several important limitations, related to generalizability, a time-limited intervention, general equilibrium effects, self-reported outcomes, and attrition. While all of these limitations are common in experimental studies of social media, they suggest caution in interpreting our results.

This project is part of the U.S. 2020 Facebook and Instagram Election Study (*49–52*). Although both the Meta researchers and academics were part of the research team, the lead academic authors had final say on the analysis plan, collaborated with Meta researchers on the code implementing the analysis plan, and had control rights over data analysis decisions and the manuscript text. Under the terms of the collaboration, Facebook could not block any results from being published. The academics were not financially compensated and the analysis plan was preregistered prior to data availability. More details of this partnership are in Appendices F and G. We have posted answers to frequently asked questions online at this website.

Theoretical Mechanisms

Prior research suggests mechanisms by which social media access could affect each of our primary outcomes.

Social media access may affect political knowledge by increasing exposure to both true and false content. Many Americans regularly get news through social media (*53*), and the great majority of political content on social media is from major news outlets (*54*). However, misinformation often spreads through social media (*55*), and a small share of users see significant amounts of false content (*54, 56, 57*).

Social media access may affect polarization by exposing users to predominantly like-minded content, due to segregated friendship networks, algorithmic filtering, and/or their own decisions (*1–3, 58*). Such users may be pulled toward more extreme attitudes or views consistent with their initial ideology. These effects may be enhanced by the prevalence of content that is extreme, sensational, or focused on negative portrayals of the out-group (*2, 59*). Polarization may also be increased by exposure to content from the *opposite* side of the political spectrum due to backlash effects or distorted perceptions (*2, 40*).

Social media access may affect perceived legitimacy of the election by exposing users to content that undermines and/or supports legitimacy, including claims of electoral fraud, unreliability of voting machines, voter suppression, foreign interference, and infringement of press freedom and free speech. Concerns about legitimacy played a particularly large role in the 2020 election, with Democrats emphasizing issues of voter access and Republicans emphasizing the possibility of voter fraud. Many also remained concerned about the possibility of foreign interference similar to what had occurred in 2016 and about infringement of press freedom and free speech (*60–64*). Claims and allegations about all of these issues were widely circulated on social media both before and after the election (*31, 64–66*).

Social media access may affect participation, including turnout and other forms such as attending protests, signing petitions, or discussing politics online. Social media provides the most important online forum for political discussions (*2*), so it could directly increase the likelihood of online participation. It could affect the likelihood of offline participation and turnout through several channels, including increasing or decreasing overall interest in the election, providing information about polling places and candidates, and simply crowding out time that would have been used for these other activities.

Finally, social media access may affect vote choice and candidate favorability. Both candidates deployed substantial resources to sway voters to their side on social media and/or encour-

age their supporters to turn out and vote. These efforts included an estimated 1.6 billion dollars spent on digital advertising, as well as various forms of unpaid content and promotion (*67*). It may have been that one campaign or the other was more effective in these efforts on net, and/or that different groups of voters were swayed in different directions. The broader body of content voters see on social media could also sway them to one side or the other.

Experimental Design

This section provides a high-level overview of the experimental design. Details are in Appendix A.

We ran two parallel experiments, with Facebook and Instagram as the respective "focal platform." For each focal platform, Meta drew a stratified random sample of users who lived in the U.S., were at least 18 years old, and had logged into their account at least once in the past month. Meta placed survey invitations in these users' focal platform news feeds from August 31 to September 12, 2020. Users who clicked on the invitations were told about the study and asked whether they were willing to deactivate their focal platform account for one week in exchange for $25 and also whether they were willing to deactivate their focal platform account for six weeks in exchange for $150. Those who were willing to deactivate were asked to complete a series of surveys, including baseline (responses collected September 8 - 21), endline (November 4 - 18), and post-endline (December 9 - 23).

Just after the baseline survey, participants were randomized into two groups: Deactivation (27 percent) and Control (73 percent). The Control group was informed that they would receive $25 if they did not log in to their focal platform for the next week, while the Deactivation group was informed that they would receive $150 if they did not log in to their focal platform for the next six weeks. Since the Deactivation condition was more expensive than Control, we allocated fewer participants to the former to increase statistical power per dollar of cost.

We included a short deactivation period for the Control group to ensure that the Control and Deactivation experiences were identical except for the deactivation length and payment amount, limiting the risk of differential attrition and experimenter demand effects. On September 23, Meta began deactivating all participants' focal platform accounts. The account deactivations were completed by the end of the day on September 23 for Instagram and September 24 for Facebook. Meta reactivated Control and Deactivation group accounts on September 30th and November 4th, respectively.

While accounts were deactivated, participants could use their Facebook and Instagram credentials to log in to other apps and services, including WhatsApp and Facebook Messenger. The experiment therefore focuses on the impact of the core Facebook and Instagram products. Participants could also choose to log in to (and thereby reactivate) the focal platform at any time. Those who did so were reminded that they would forfeit their deactivation payments, but they were asked to remain in the study and complete the remaining surveys. In addition to the deactivation incentives, participants received base payments of $5 for completing the baseline survey and $20 each for the endline and post-endline surveys.

In addition to variables drawn from the baseline, endline, and post-endline surveys, our analysis incorporates a number of directly measured variables. First, we matched participants to state voting records to produce a validated measure of voter turnout. Second, we matched participants to public records of campaign donations. Third, a subset of participants opted in to passive tracking that directly recorded their use of news and social media apps and websites in exchange for an additional payment. Finally, we matched all participants in our primary analysis sample to Meta platform data, including time spent on the platforms.

Empirical Strategy

Our pre-analysis plan was registered with the Open Science Foundation on September 22, 2020.[1] It specified the primary analysis sample, the target population for creating sample weights, the primary and secondary outcomes, the subgroup analyses, the estimating equation, and the handling of missing data. It also included shells of key tables and figures. It specified that we would base inference on two-sided tests and that we would use sharpened False Discovery Rate (FDR) adjusted q-values (*68*) to control for multiple hypothesis testing, with $q < 0.05$ as our measure of statistical significance for all tests. We did not substantively deviate from the pre-analysis plan. Appendix H describes clarifications and minor modifications, mainly driven by changes in data availability. At the request of the editor, we characterize our null results using equivalence bounds, which are identical to 90 percent confidence intervals without adjustments for multiple hypothesis testing. We note that in some cases, the equivalence bounds both have the same sign even though the effect is not statistically significant at our preregistered thresholds.

Our primary analysis samples are limited to participants who used the focal platform more than 15 minutes per day at baseline. We weight the samples to be representative of U.S. focal platform users on race, political party, education, and (among those with more than 15 minutes of daily use) baseline account activity. See Appendix A.5 for additional details.

We use an instrumental variables regression to estimate the causal effect of deactivation while accounting for imperfect compliance with deactivation. In Appendix D.1, we report intent-to-treat effects based on ordinary least squares regressions of outcomes on an indicator for assignment to treatment. We define D_i as a measure of participant i's deactivation compliance, $\boldsymbol{X}_i$ as a vector of controls, ν_s as a vector of randomization stratum indicators, and Y_i as an outcome. The estimating equation is

[1]The PAP is publicly available from http://www.osf.io/h3c6w/.

$$Y_i = \tau D_i + \rho \boldsymbol{X}_i + \nu_s + \varepsilon_i, \quad (1)$$

where we instrument for D_i with a Deactivation group indicator variable T_i.

Appendix B includes more details on variable construction. As pre-specified, the controls $\boldsymbol{X}_i$ are the variables selected in a lasso regression of Y_i on the baseline value of Y_i (if available) and a vector of demographics and baseline survey variables. We use the same vector $\boldsymbol{X}_i$ in all subgroup analyses for a given outcome.

The compliance measure is defined as $D_i = (\bar{U}_C - U_i)/\bar{U}_C$, where U_i is the share of days that i used the platform (defined by viewing five or more pieces of content) during the September 30 - November 3 treatment period and $\bar{U}_C$ is the Control group average. Thus, $D_i = 1$ for participants who never use the platform, $D_i = 0$ for participants with usage equal to the Control group average, and τ measures the local average treatment effect of never using the platform instead of using the Control group average, for people induced to deactivate by the $150 payment.

We emphasize a number of important limitations that readers should keep in mind in interpreting the results below. Similar caveats also apply to prior experimental studies of social media including (*5*), (*37*), and (*38*). First, readers should be cautious about generalizing beyond our specific sample and time period. Our estimates are only directly informative about the set of people who agreed to participate in the study and were willing to deactivate their accounts for the payments we offered, and we do not know what the effects might have been in a different year or outside of an election period. However, we do think our results can inform readers' priors about the potential effects of social media in the final weeks of high-profile national elections.

Second, we can only estimate the effect of five weeks of individual-level deactivation. Longer-term deactivation could have different effects, and deactivating many users or elimi-

nating Facebook or Instagram altogether could have broader "general equilibrium" effects on how people get information and what media outlets cover. Third, although we include a number of directly measured outcomes, many key outcome variables are self-reported, raising the possibility of various forms of measurement error. Fourth, although we designed our experiment to minimize the possibility of experimenter demand effects (e.g., requiring both Control and Deactivation participants to deactivate their accounts for at least some time), and although past work (*5, 69*) implies that experimenter demand effects in our context may be limited, participants were aware that they were part of an experiment, and we cannot rule out the possibility that this influenced their behavior.

Sample, Compliance, and Substitution

Sample characteristics. We summarize the characteristics of our sample in Appendix C. On Facebook, a total of 10.6 million users were invited to the study. 673,388 clicked the invitation, of whom 52,821 consented to participate and were willing to deactivate. Of these, 19,857 completed the baseline survey, could be linked to platform data, and had at least 15 minutes of baseline use per day. This final group is our "primary analysis sample." Of the primary analysis sample, 5,691 people opted into passive tracking. On Instagram, the analogous numbers are 2.6 million invites, 319,271 clicks, 51,955 consents, 15,585 participants in the primary analysis sample, and 3,822 people in passive tracking.

Relative to the broader target populations of Facebook and Instagram users, our sample over-represents users who are more liberal and more engaged with civic content. Our weighting strategy addresses selection on these variables. In Appendix Tables S6 and S7, we show that differential selection on these characteristics is mainly driven by the users who select into our screening survey at the initial stage, and that characteristics of the sample are relatively stable from then on.

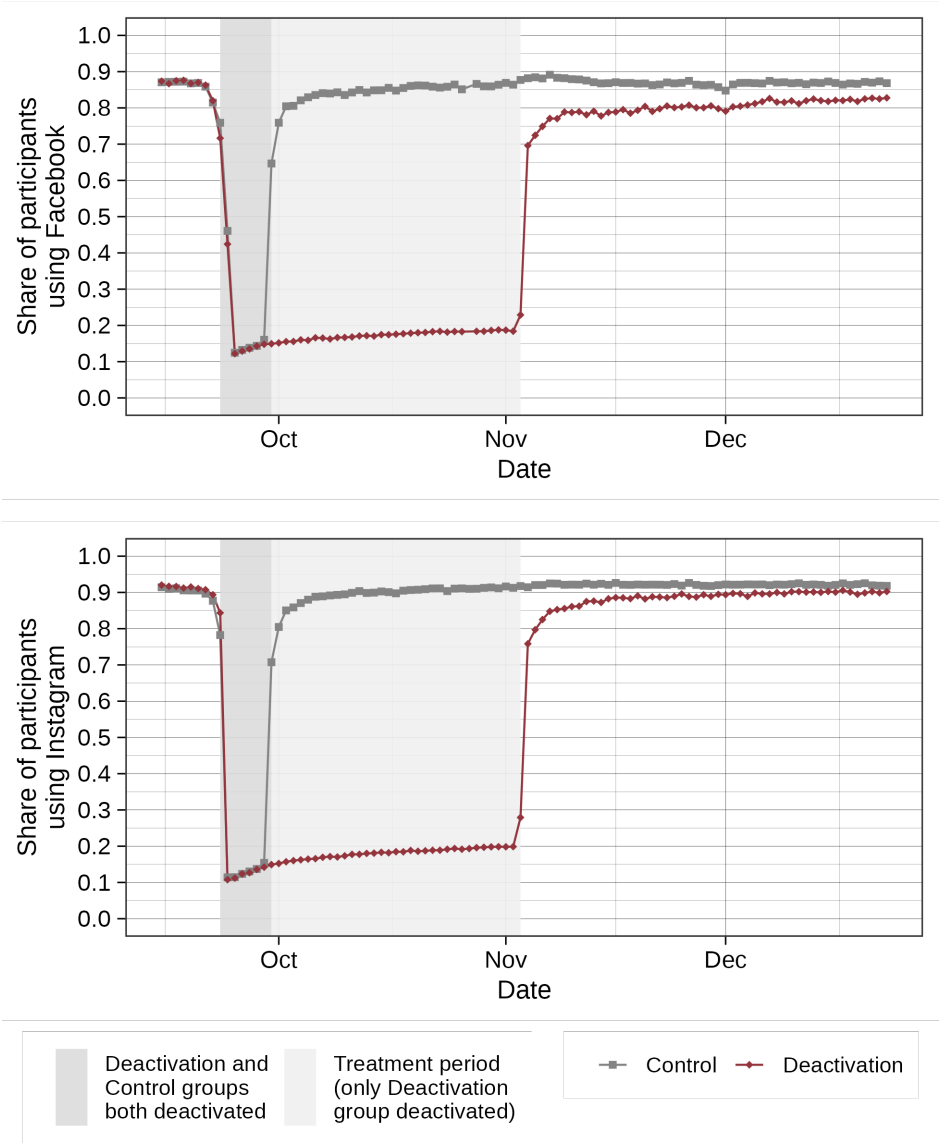
The degree of selection into our experiment is comparable to that in the prior literature. In (*5*), recruitment ads were shown to 1.9 million users; 2,897 completed baseline and were randomized. In (*38*), recruitment ads were shown to 365,599 users; 556 completed baseline and were randomized. In (*37*), recruitment ads were shown to a large set of Texas A&M undergraduates, with the exact number who received them not reported; of the 1,929 individuals who responded, 167 completed baseline and were randomized.

Balance and attrition. The Deactivation and Control groups were balanced on observable characteristics at both baseline and endline; see Appendix Tables S12 and S13. In our Facebook primary analysis sample, 91 percent of the Deactivation group and 89 percent of the Control group completed the endline survey. In the Instagram sample, the corresponding shares are 88 percent and 86 percent. Our attrition rates of 9–14 percent compare favorably to prior work—e.g., they are similar to (*37*), lower than (*38*), and lower than the mean of 96 field experiments published in top economics journals that were surveyed by (*70*).

The roughly 2 percentage point difference between attrition rates in Deactivation and Control is also comparable to prior work (*70*). Given our large sample, these differences are statistically significant for both samples; see Appendix Table S14. There were also differences in the timing of response, with Deactivation participants generally completing the endline survey earlier than Control participants; see Appendix Figure S2. While this attrition is an important potential concern, below we discuss a range of evidence suggesting that it is unlikely to be a source of substantial bias in our estimates.

Compliance. Figure 1 presents the share of participants who used Facebook and Instagram (defined as logging in and seeing at least five pieces of content) for each day of the experiment. Before the experiment, about 90 percent of users in the Deactivation and Control groups accessed the focal platform on a given day. During the period when both groups were being paid

Figure 1: Share of Participants Using Facebook and Instagram During Study Period



Note: This figure presents the share of Deactivation and Control groups that used Facebook and Instagram on each day. "Use" is defined as logging in and seeing five or more pieces of content. The dark grey shaded area indicates the Control group's 7-day deactivation period, while the light grey shaded area indicates the Deactivation group's 35-day additional deactivation period. We exclude Facebook use data from October 27th due to a logging error.

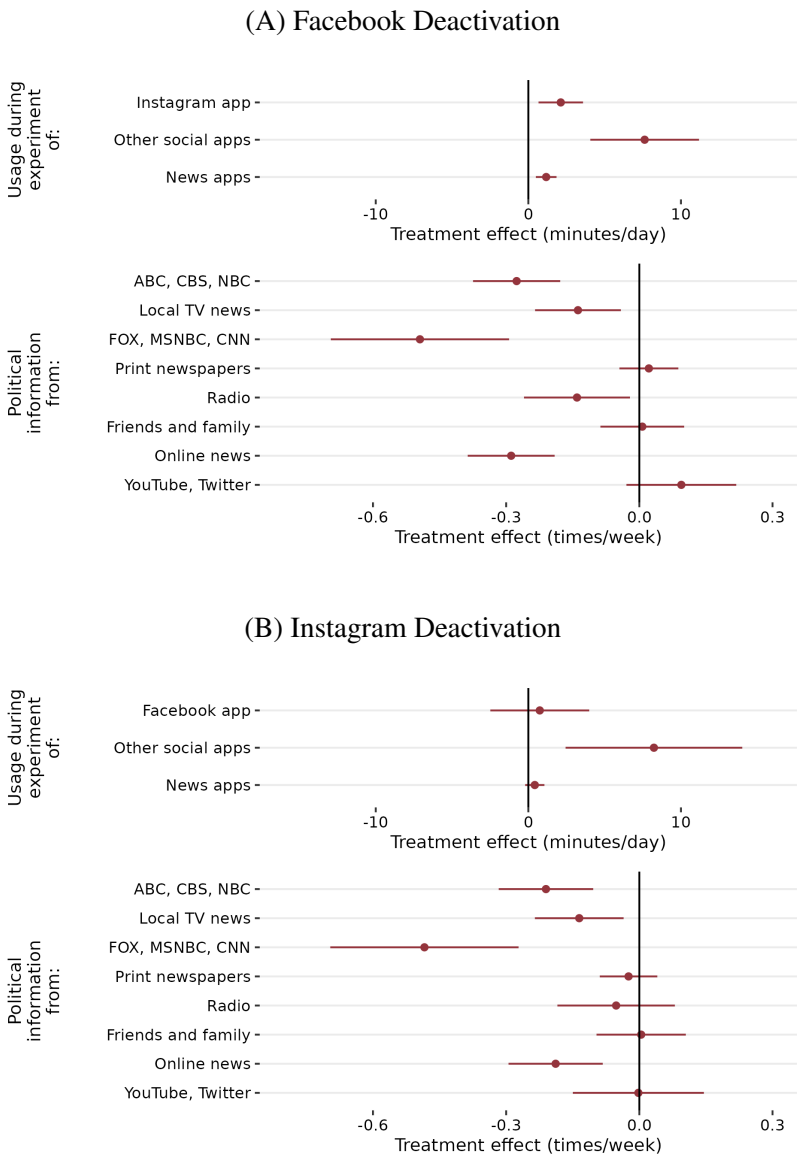
to avoid logging in and the deactivation process had been completed for all users (September 25-29 for Facebook and September 24-29 for Instagram), the daily usage rate was roughly 10 to 15 percent. From September 30 - November 3, the five-week "treatment period" when only the Deactivation group was being paid to avoid logging in, that group's daily usage rate was roughly 15 to 20 percent, while the Control group's usage rate returned close to baseline levels.

Thus, the experiment reduced the Deactivation group's daily usage rate by about $(0.9 - 0.15)/0.9 \approx 83\%$ relative to Control over the treatment period. This illustrates the first stage of the instrumental variables estimator in equation (1). Consistent with results in (*5*) and (*71*), the Deactivation group was still less likely to use the platforms at the end of December, almost two months after the deactivation incentive had ended. In Appendix Figure S5 and Appendix Table S29, we show this amounted to reductions (relative to Control) of 24 percent and 14 percent in average time spent over the post-deactivation period for Facebook and Instagram, respectively.

Substitution. The treatment effects we estimate below are the combined effect of reducing Facebook or Instagram use *and* reallocating that time to other activities. How that time is reallocated is a key determinant of the effects (*72*). For example, if participants reallocated most of their social media time to consuming high-quality news sources that generally provide accurate information, the impacts could be quite different than if they allocated most of that time to less accurate sources or to activities unrelated to news or politics.

Figure 2 presents the effects of Facebook and Instagram deactivation on use of substitutes. The first three rows of each panel show effects on time spent on other mobile apps, as directly measured among participants who opted into passive tracking. As a point of reference, the Facebook and Instagram Control group participants in the passive tracking sample spent about 43 and 16 minutes per day, respectively, on the Facebook and Instagram mobile apps during the treatment period. Facebook deactivation increased time spent on Instagram, other social media

Figure 2: Effects of Facebook and Instagram Deactivation on Use of Substitutes



Note: This figure presents local average treatment effects of Facebook (Panel A) or Instagram (Panel B) deactivation estimated using equation (1). Facebook / Instagram app minutes, other news apps minutes, and minutes on other social apps are measured only for participants who opted into passive tracking of apps. The bottom rows in each panel use answers to the question, "How often in the past week have you gotten political information from the following sources?" Responses were coded as "every day" = 7, "several times" = 3, "once" = 1, and "never" = 0. The horizontal lines represent 95 percent confidence intervals.

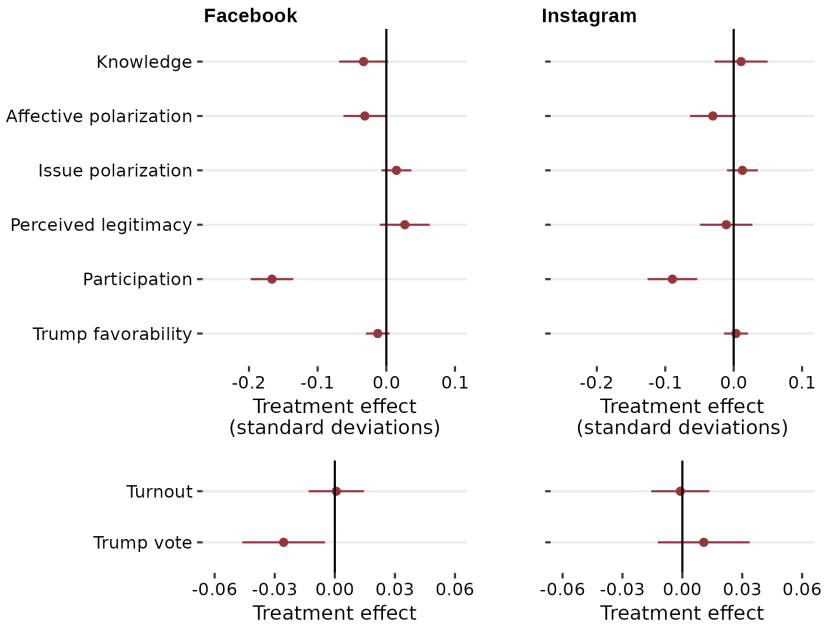
apps (such as YouTube, Twitter, and Snapchat), and news apps (such as the New York Times and Fox News), by point estimates of about 2, 8, and 1 minutes per day, respectively. Together, these point estimates sum to about 26 percent of the passive tracking sample Control group's average Facebook mobile app use the during the treatment period, with the remaining time going to activities we did not observe. Instagram deactivation caused almost no substitution to Facebook or news apps, but it increased time spent on other social media apps by a point estimate of 8 minutes per day. Together, the point estimates sum to about 57 percent of the passive tracking sample Control group's average Instagram mobile app use.

The bottom eight rows of each panel show that both Facebook and Instagram deactivation reduced the amount of political information that participants reported getting from online news outlets, as well as from network, cable, and local TV news outlets. Facebook deactivation also reduced political information from radio. In total, across all sources reported, the Facebook and Instagram Deactivation groups reported getting political information just over one time per week less. This is a reduction of about 4 percent relative to the Control group average of 26 times per week. One explanation is that deactivation reduces the amount of content from these outlets that people receive through Facebook or Instagram, and this is not outweighed by any increase in direct access. Other possible explanations are that deactivation reduced attention to political content on these other media or that these self-reports are inaccurate (*73*).

Effects on Primary and Secondary Outcomes

Figure 3 presents the effects of deactivation on our eight pre-specified primary outcome variables. The first six variables are standardized into units of standard deviations within the Control groups. The bottom two variables (turnout and Trump vote) are reported in their original units. Appendix D.1 presents point estimates, p values, and equivalence bounds for all primary outcomes.

Figure 3: Effects of Facebook and Instagram Deactivation on Primary Outcomes

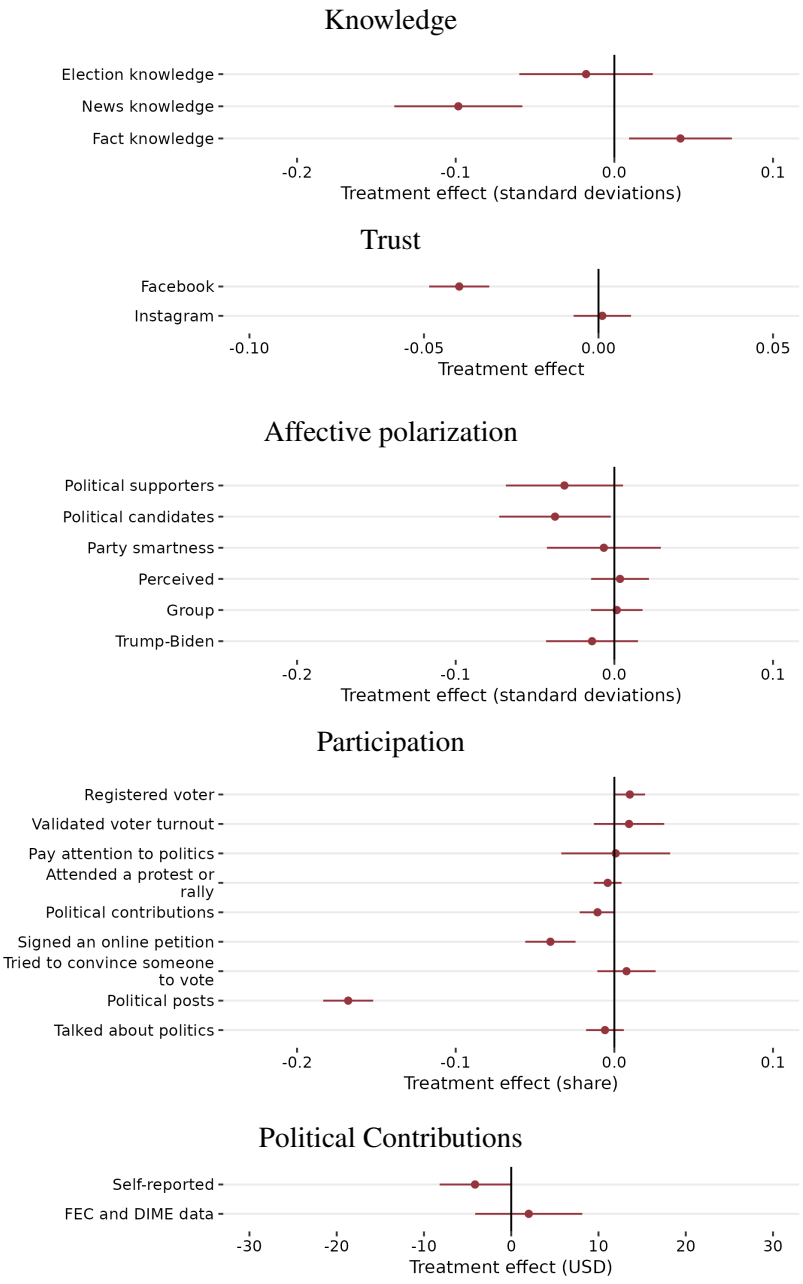


Note: This figure presents local average treatment effects of Facebook and Instagram deactivation estimated using equation (1). The horizontal lines represent 95 percent confidence intervals.

Figure 4 presents the effects of Facebook deactivation on selected outcomes that we labeled as "secondary" in the pre-analysis plan but are useful in interpreting the results in Figure 3. Appendix Figure S3 presents the analogous results for Instagram. Appendix D.2 presents point estimates and p values for the full set of preregistered secondary outcomes.

Knowledge. The first primary outcome is political knowledge. The *knowledge* variable is the average of standardized scores on three sets of factual questions: (i) *election knowledge* (knowledge of candidates' policy positions); (ii) *news knowledge* (correctly distinguishing recent news events from plausible placebo events that had not happened); and (iii) *fact knowledge* (correctly distinguishing true statements from misinformation that was circulating about topics such as COVID-19 and fraudulent ballots). These questions were based on a set of true and false news stories that circulated widely during the period in which our treatments took place.

Figure 4: Effects of Facebook Deactivation on Selected Secondary Outcomes



Note: This figure presents local average treatment effects of Facebook deactivation estimated using equation (1). "Trust" uses answers to the question, "How much do you think political information from each of these sources can be trusted?" Responses were coded as "not at all" = 0, "a little" = 0.25, "a moderate amount" = 0.5, "a lot" = 0.75, and "a great deal" = 1. The horizontal lines represent 95 percent confidence intervals.

The effects of both Facebook and Instagram deactivation on our overall knowledge index are small and insignificant. For Facebook, the point estimate is that deactivation slightly reduced *knowledge* by 0.033 standard deviations ($p = 0.069$, $q = 0.190$, equivalence bounds $= -0.063, -0.003$). For Instagram, the point estimate is closer to zero (equivalence bounds $= -0.022, 0.043$).

This overall *knowledge* index masks two partially offsetting components, as shown in the first panel of Figure 4. First, Facebook deactivation decreased *news knowledge* by a point estimate of 0.098 standard deviations ($p < 0.01$, $q < 0.01$). This is directionally consistent with the reduction in self-reported news consumption from Figure 2, and the effect size is similar to that in (*5*). Second, the (statistically insignificant) point estimate is that Facebook deactivation increased *fact knowledge* by 0.042 standard deviations ($p = 0.012$, $q = 0.132$). As a point of comparison for these magnitudes, college graduates in the Control groups scored 0.54 and 0.57 standard deviations higher than non-college graduates on *news knowledge* and *fact knowledge*, respectively.

Auxiliary analyses in Appendix Figure S7 show effects on the individual survey questions used to construct these knowledge indices. Within the *news knowledge* index, Facebook deactivation decreased correct beliefs in three events that did happen ((i) "a militia's plot to kidnap Michigan governor Gretchen Whitmer was foiled by undercover agents," (ii) "Pope Francis voiced support for same-sex civil unions," and (iii) "Amy Coney Barrett became the newest Supreme Court justice") and increased incorrect beliefs in one event that did not ("Donald Trump announced that he would stop holding public rallies out of concern for COVID-related risks"). Within the *fact knowledge* index, Facebook deactivation did not affect beliefs in any true statements but decreased belief in three false statements: (i) "evidence found on Hunter Biden's laptop proves Joe Biden took bribes from foreign powers," (ii) "millions of fraudulent ballots were cast in the 2020 presidential election," and (iii) "Joe Biden is a pedophile." This

evidence is consistent with Facebook access increasing belief in misinformation.

The second panel of Figure 4 shows that Facebook deactivation reduced trust in political information from Facebook by a point estimate of 0.040 standard deviations ($p < 0.01$, $q < 0.01$), without affecting trust in Instagram. Appendix Figure S3 similarly shows that Instagram deactivation reduced trust in information from Instagram without affecting trust in Facebook. One potential explanation is that time away from a platform made users more aware of the amount of low-quality or inaccurate information to which they had been exposed. The effect echoes the finding in (*5*) that Facebook deactivation reduced post-experiment Facebook use, and in (*74*) that Fox News viewers who had been given incentives to watch CNN for a month reported lower trust in Fox News. Appendix Tables S21 and S22 show that deactivation did not impact trust in other news sources.

Polarization. We construct two measures of political polarization: *affective polarization* and *issue polarization*. For both variables, we exclude Independents who report that they do not lean toward one party or the other.

Affective polarization is the standardized average of three underlying variables: (i) *political supporters polarization* (the difference in participants' favorability toward people who support their own party vs. the other party); (ii) *political candidates polarization* (the difference in participants' favorability toward people running for office in their own party vs. the other party); and (iii) *party smartness polarization* (the difference in participants' perceived smartness of people in their own party vs. people in the other party). Neither Facebook nor Instagram deactivation had a significant effect. The (statistically insignificant) point estimates are that both Facebook and Instagram deactivation reduced *affective polarization* by 0.031 and 0.030 standard deviations (Facebook: $p = 0.049$, $q = 0.189$, equivalence bounds $= -0.06, -0.005$; Instagram: $p = 0.074$, $q = 0.190$, equivalence bounds $= -0.06, -0.002$).

The third panel of Figure 4 presents the effects of Facebook deactivation on the three components of the *affective polarization* index, plus the other three preregistered affective polarization secondary outcomes: perceived polarization, polarization in views towards groups such as immigrants and minorities, and polarization in feeling thermometer ratings for Trump and Biden. All six effects are statistically indistinguishable from zero after adjusting for multiple hypothesis testing. The largest point estimates are for *political candidate polarization* and *political supporters polarization*.

The *issue polarization* variable is an index of eight political opinions (on immigration, repeal of Obamacare, unemployment benefits, mask requirements, foreign policy, policing, racial justice, and gender relations), with the signs of the variables adjusted so that the difference between the own-party and other-party averages is positive. These questions were chosen to focus on issues that were prominent during the study period. Neither Facebook nor Instagram deactivation significantly affected *issue polarization*, and the equivalence bounds rule out effects of ± 0.04 standard deviations.

As a point of comparison for these magnitudes, (*5*) find that Facebook deactivation reduced an overall index of political polarization prior to the 2018 midterm elections. This includes a statistically insignificant reduction of 0.06 standard deviations in a measure of affective polarization, and a significant reduction of 0.10 standard deviations in a measure of issue polarization. One possible explanation for the difference in effects on issue polarization is that our study took place during a presidential election, where the environment was saturated with political information and opinion from many sources outside of social media. Another possible explanation is that the set of specific issues we focus on here may have produced different responses. As another comparison point, (*26*) estimate that affective polarization has grown by an average of 0.021 standard deviations *per year* since 1978.

Perceived legitimacy of election. The *perceived legitimacy* variable is an index of agreement with six statements: (i) elections are free from foreign influence, (ii) all adult citizens have equal opportunity to vote, (iii) elections are conducted without fraud, (iv) government does not interfere with journalists, (v) government protects individuals' right to engage in unpopular speech, and (vi) voters are knowledgeable about candidates and issues. Neither Facebook nor Instagram deactivation had a significant effect, and the equivalence bounds rule out effects of ± 0.06 standard deviations.

These null results are important in light of the prominent challenges to election legitimacy that took place around the 2020 election, including Trump's emphasis on potential fraud in the run-up to the election and the "Stop the Steal" movement that followed it. Of the sub-components of our election legitimacy index, one might expect (iii) to be most related to these events. Appendix Tables S21 and S22 show that effects of deactivation on this component are also small and insignificant. We show below that the same is true for effects on our post-endline survey in December 2020. We also discuss below the separate effects on Republicans and Democrats.

Political participation and turnout. The *participation* variable is the sum of indicators for whether a participant reported doing the following six activities: (i) attended a protest or rally, (ii) contributed money to a political candidate or organization, (iii) signed an online petition, (iv) tried to convince someone how to vote (online or in-person), (v) wrote and posted political messages online, and (vi) talked about politics with someone they know. *Turnout* is an indicator for whether the participant reported voting.

Facebook and Instagram deactivation significantly reduced *participation*, by point estimates of 0.167 and 0.090 standard deviations, respectively ($p < 0.01$, $q < 0.01$ for both variables). Neither Facebook nor Instagram deactivation significantly affected *turnout*, and the equivalence

bounds rule out effects larger than about 1.3 percentage points in either direction.

The fourth panel of Figure 4 presents effects of Facebook deactivation on the six components of *participation*, plus the five other preregistered participation secondary outcomes. The effects of Facebook on *participation* are mainly driven by online activities: posting and signing petitions online. This panel also shows that the effect on turnout is also close to zero and statistically insignificant when we use the directly measured *validated voter turnout* variable. These results are similar for Instagram; see Appendix Figure S3.

The final panel of Figure 4 presents effects on political contributions measured in dollars, using both our survey-based outcome and our directly-measured outcome based on administrative contribution records. Both effects are statistically insignificant after adjusting for multiple hypothesis testing.

Candidate evaluations and vote choice. The *Trump favorability* variable is the average of two components: (i) Trump approval ratings and (ii) the difference between Trump and Biden feeling thermometer ratings. The *Trump vote* variable is defined as 1 for people who reported voting for Trump, -1 for people who reported voting for Biden, and 0 for those who didn't vote or voted for some other candidate. To illustrate the units, a 0.03 unit reduction in *Trump vote* would result if Facebook deactivation caused 1.5 percent of Trump voters to switch to Biden, or if it caused 3 percent of Trump voters to not vote.

Neither Facebook nor Instagram deactivation significantly affected *Trump favorability*. The point estimate is negative for Facebook (equivalence bounds $= -0.027, 0.002$) and very close to zero for Instagram (equivalence bounds $= -0.011, 0.018$). The point estimate for the effect of Facebook deactivation on *Trump vote* is a reduction of 0.026 units ($p = 0.015$, $q = 0.076$, equivalence bounds $= -0.043, -0.008$). This effect is not significant at our preregistered threshold of $q < 0.05$. Instagram deactivation did not significantly affect *Trump vote* (equivalence bounds

$= -0.009, 0.030$). Appendix D.4 presents additional analyses of the voting data. As discussed in Appendix G.2, the samples were far too small for the experiment itself to have changed any actual election outcomes.

Several additional analyses help assess whether an effect of Facebook deactivation on Trump vote share would be plausible. First, some people did change their minds over the study period: about 20 percent of Control group participants report voting differently at endline than they had reported intending to vote at baseline (see Appendix Table S24). Second, several (statistically insignificant) point estimates on other variables are directionally consistent with an effect on vote choice: deactivation (statistically insignificantly) decreased *Trump favorability* as described above, (statistically insignificantly) decreased turnout among Republicans and increased turnout among Democrats (see Appendix Figure S21), and (statistically insignificantly) decreased three secondary outcomes reported in Appendix D.2: pro-Republican affect, pro-Republican issue positions, and votes for Republicans in state-level elections. However, the fact that any effects on correlated outcomes are small is an important note of caution. [2]

Post-endline effects. There are no statistically significant effects of Facebook or Instagram deactivation on the affective polarization, perceived legitimacy, and Trump favorability questions on the post-endline survey; see Appendix Figure S6. This is not surprising, given that there were no statistically significant effects on these outcomes at endline and the Deactivation groups used Facebook and Instagram almost as much as the Control groups for the month before the post-endline survey.

[2]For example, a one standard deviation increase in *Trump favorability* is associated with a 0.6 unit increase in *Trump vote* in the Control group. Since deactivation reduced *Trump favorability* by a point estimate of 0.013 standard deviations (with a 95 percent confidence lower bound of 0.030 standard deviations), one might thus expect *Trump vote* to decrease by $0.013 \times 0.6 \approx 0.008$ units (with a 95 percent confidence lower bound of 0.018 units), which is closer to zero than the actual point estimate of 0.026 units.

Attrition and robustness. Earlier, we reported that 9–14 percent of the participants in our primary analysis sample failed to complete the endline survey, and that these attrition rates were somewhat higher for the Deactivation groups than for the Control groups. As shown formally by (*70*), the key question for assessing the potential for attrition to introduce bias is not whether it is differential *per se*, but instead the extent to which it generates different distributions of potential outcomes across the treatment conditions. We present several pieces of evidence that together suggest such bias may be limited.

First, we apply the test proposed by (*70*) to evaluate potential bias using the distribution of baseline outcomes among attritters and non-attritters. This test, which is presented in Appendix Tables S25 and S26, fails to reject the null hypothesis that the distribution of potential outcomes remains balanced post-attrition for 15 of the 16 primary outcomes across the two experiments. The one rejection is for affective polarization in the Facebook sample. The baseline means of affective polarization are balanced in the endline sample, meaning that there is no evidence of systematic distortion in our endline treatment effect estimates, but the baseline means are less balanced among attritters.

Second, we present additional evidence on the extent to which attrition is related to observed determinants of outcomes. Appendix Table S13 shows that the Deactivation and Control groups are balanced on observables at endline. Consistent with this, Appendix Figure S10 shows that the estimates depend little on the choice of control variables $\boldsymbol{X}_i$. Appendix Figure S15 shows that treatment effects on endline outcomes predicted using baseline observables are all indistinguishable from zero. Appendix Tables S27 and S28 present bias tests recommended by (*75*) which suggest minimal bias from attrition under the assumption that the distribution of unobserved outcome determinants is similar to the distribution of observables.

Third, in Appendix Table S29, we use our directly-measured outcomes (which are available for both attritters and non-attritters) to compare the results for the full sample vs. the post-

attrition sample. This provides a direct test of bias due to attrition for these outcomes. In all cases, the full and post-attrition samples yield very similar results, with the differences between the two in all cases less than one standard error. This shows that attrition is not a significant source of bias in the treatment effects for these outcomes.

Finally, we construct Lee (*76*) bounds for the worst-case bias that could have been produced by attrition; see Appendix Tables S30 and S31. These show how the estimates would change if those who attrited were maximally selected on *unobserved* determinants of outcomes. Following (*76*), we delete the highest or lowest values in the distribution of Deactivation group outcomes until the resulting response rate equals the Control group's response rate; we also "tighten" the estimates using baseline characteristics. The Lee bounds exclude zero for three primary outcomes: *participation* for both Facebook and Instagram and *Trump vote* for Facebook.

In Appendix Figures S10, S11, and S12, we present effects on our primary outcomes under several alternative specifications: (i) weighting observations equally; (ii) controlling for party by date of survey response to address possible underlying time trends in partisan views in the days after the election;[3] (iii) excluding the lasso-selected control variables $\boldsymbol{X}_i$ from the estimating equation; (iv) excluding the 8 percent of participants in the Facebook sample and 33 percent of participants in the Instagram sample who reported having multiple accounts; and (v) using alternative weights that vary the threshold for winsorization and omit individual variables from the set used to calculate the weights one at a time. None of these robustness checks substantially changes the qualitative or quantitative pattern of our estimates.

Subgroup analysis. As pre-specified, we estimate heterogeneous effects using four primary moderators (political party, baseline use, and indicators for undecided voters and for voters

[3]From November 4 - 16, *perceived legitimacy* dropped substantially among Republicans while increasing substantially among Democrats; see Appendix Figures S13 and S14.

who identify as Black / Hispanic) and five secondary moderators (age, gender, education, urban status, and an indicator for swing state residence). These results are reported in Appendix E with the relevant confidence intervals for each subgroup.

The effects of Facebook deactivation on perceived legitimacy and participation are larger for the above-median use group. There are no other significant differences by baseline use. The effects on undecided voters and Black/Hispanic voters are generally indistinguishable from both the full sample estimates and from zero. Furthermore, there are generally no significant differences in primary effects within each of these subgroups (minority vs. nonminority voters; undecided vs. decided voters).

Effects for political party subgroups are generally similar to each other, with several exceptions. First, Facebook deactivation significantly reduced our overall index of *knowledge* among Democrats. Second, Facebook deactivation increased *issue polarization* among strong Democrats and (insignificantly) decreased it among strong Republicans; see Appendix Figure S21. This could reflect Facebook deactivation making the issue views of both Republicans and Democrats more pro-Democratic. Third, both Facebook and Instagram deactivation initially appear to have increased perceived legitimacy among Republicans, but this result does not survive controlling for the timing of survey response; see Appendix Figures S21 and S22.

In terms of the secondary moderators, most differences are not significant. One notable pattern is that there are several subgroups (above-median age, men, people with college degrees, and rural residents) in which Facebook deactivation had somewhat larger and more significant negative effects on *affective polarization*. Appendix E provides additional discussion of these and other differences, including relevant confidence intervals for each subgroup.

Discussion

Our findings paint a nuanced picture of the way Facebook and Instagram influenced attitudes, beliefs, and behavior in the 2020 election.

We replicate in a new context and a larger sample the prior finding by (*5*) that access to Facebook leads users to be more informed about events in the news. We are also the first to present large-scale experimental evidence suggesting that access to Facebook may increase belief in misinformation. Together, these findings are consistent with the fact that both trustworthy news and misinformation circulate widely on the platform, and they highlight the importance of continuing efforts to decrease exposure to the latter.

We find no significant effects of social media access on polarization. Our point estimates are consistent with both Facebook and Instagram access slightly increasing affective polarization, but the effects are small relative to the long-term trend in affective polarization. We find no significant effects of either Facebook or Instagram access on issue polarization. This suggests that if Facebook access contributes to political polarization, the effect is either small or accumulates over a longer period than five weeks.

Our estimated effect of Facebook access on self-reported vote choice is large enough to be meaningful in a close election, although it was not significant at our preregistered threshold. Any effect would be consistent with the Trump campaign using Facebook more effectively than the Biden campaign. It could also reflect a net effect of the posts from users, pages, and groups to which users were exposed. We reiterate that our vote choice estimate applies to the specific population that selected into our experiment and that it cannot be extrapolated to the broader population without strong assumptions. We also reiterate that our sample was far too small for the experiment itself to have changed any actual election outcomes.

Our results provide the first ever large-scale evidence on the political impact of access to

Instagram. Aside from a reduction in online participation, we find no significant impacts of Instagram on any other primary outcomes. This is true even among younger users, and it suggests that despite Instagram's rapid growth, Facebook likely remains the platform with the largest impacts on political outcomes.

We find precisely estimated zero effects of both Facebook and Instagram access on the perceived legitimacy of the electoral process, including perceptions of electoral fraud. Of course, we do not capture the effect of content shared after election day because our deactivation period had ended.

Finally, we find no significant effects of either Facebook or Instagram access on turnout, with confidence intervals that rule out moderate effects. This is true for turnout measured with our survey as well as validated turnout from the voter file.

All of these findings are subject to the important caveats described above, including generalizability, a time-limited intervention, general equilibrium effects, self-reported outcomes, and attrition. Notwithstanding its limitations, we believe that this study can usefully inform and constrain the discussion of the effects of social media on American democracy.

Data, Materials, and Code Availability

De-identified data and analysis code from this study will be stored in the Social Media Archive (SOMAR) housed by ICPSR. This will be accessible for university research approved by the Institutional Review Board (IRB) related to elections or for the purpose of validating the results of this study. ICPSR will handle and vet all applications requesting access to this data.

Ethics and Competing Interests

Some authors are employed by Meta. None of the academic researchers nor their institutions received financial or any other compensation from Meta for their participation in the project.

We have adhered to all relevant ethical guidelines. The overall project was evaluated and approved by the NORC Institutional Review Board (IRB). Academic researchers coordinated with their specific university IRBs to ensure they followed regulations concerning human subject research when analyzing data collected by NORC and Meta as well as when authoring papers based on the results. Additionally, the research group was provided ethical counsel by the independent company Ethical Resolve to inform the study designs.

For more information, see Appendices F and G.

References and Notes

1. C. R. Sunstein, *#Republic: Divided Democracy in the Age of Social Media* (Princeton University Press, 2017).

2. J. E. Settle, *Frenemies: How Social Media Polarizes America*.

3. P. Barberá, *Social media and democracy: The state of the field, prospects for reform*, N. Persily, J. A. Tucker, eds., SSRC Anxieties of Democracy (Cambridge University Press, 2020), p. 34–55.

4. E. Zhuravskaya, M. Petrova, R. Enikolopov, Political effects of the internet and social media, *Annual Review of Economics* **12**, 415-438 (2020).

5. H. Allcott, L. Braghieri, S. Eichmeyer, M. Gentzkow, The Welfare Effects of Social Media, *American Economic Review* **110**, 629-76 (2020).

6. J. J. Van Bavel, *et al.*, How social media shapes polarization, *Science and Society* **25**, 913-916 (2021).

7. Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (Yale University Press, 2008).

8. R. Fletcher, R. K. Nielsen, Are people incidentally exposed to news on social media? A comparative analysis, *New Media and Society* **20**, 2450-2468 (2018).

9. K. Munger, P. J. Egan, J. Nagler, J. Ronen, J. Tucker, Political Knowledge and Misinformation in the Era of Social Media: Evidence From the 2015 UK Election, *British Journal of Political Science* **52**, 107–127 (2022).

10. C. Silverman, This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook, Available at https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook (2016).

11. L. McIntyre, *Post-truth* (MIT Press, 2018).

12. S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, *Science* **359**, 1146-1151 (2018).

13. A. M. Guess, B. A. Lyons, *Social Media and Democracy: The State of the Field, Prospects for Reform*, N. Persily, J. A. Tucker, eds., SSRC Anxieties of Democracy (Cambridge University Press, 2020), p. 10–33.

14. R. M. Bond, *et al.*, A 61-million-person experiment in social influence and political mobilization, *Nature* **289**, 295-298 (2012).

15. J. J. Jones, R. M. Bond, E. Bakshy, D. Eckles, J. H. Fowler, Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 U.S. presidential election, *PLOS ONE* **12**, 1-9 (2017).

16. A. Ravel, A new kind of voter suppression in modern elections (2019).

17. S. Overton, State power to regulate social media companies to prevent voter suppression (2020).

18. J. A. Tucker, Y. Theocharis, M. E. Roberts, P. Barberá, From Liberation to Turmoil: Social Media And Democracy, *Journal of Democracy* **28**, 46-59 (2017).

19. S. McKay, C. Tenove, Disinformation as a threat to deliberative democracy, *Political Research Quarterly* **74**, 703-717 (2021).

20. H. J. Parkinson, Click and elect: How fake news helped Donald Trump win a real election, Available at https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies (2016).

21. M. Read, Donald Trump won because of Facebook, Available at https://nymag.com/intelligencer/2016/11/donald-trump-won-because-of-facebook.html (2016).

22. J. Clayton, Social media: Is it really biased against US Republicans?, Available at https://www.bbc.com/news/technology-54698186 (2020).

23. Pew Research Center, Most Americans think social media sites censor political viewpoints, Available at https://www.pewresearch.org/internet/2020/08/19/most-americans-think-social-media-sites-censor-political-viewpoints/ (2020).

24. K. Taylor, Instagram is gen Z's go-to source of political news — and it's already having an impact on the 2020 election, Available at https://www.businessinsider.com/gen-z-gets-its-political-news-from-instagram-accounts-2019-6 (2019).

25. H. Murphy, D. Sevastopulo, Why US politicians are turning to Instagram ahead of 2020 election, Available at https://www.ft.com/content/737d2428-2fdf-11e9-ba00-0251022932c8 (2019).

26. L. Boxell, M. Gentzkow, J. M. Shapiro, Cross-Country Trends in Affective Polarization, *The Review of Economics and Statistics* pp. 1–60 (2022).

27. C. Dewey, Facebook fake-news writer: 'I think Donald Trump is in the White House because of me', Available at https://www.washingtonpost.com/news/the-intersect/wp/2016/11/17/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me (2016).

28. M. Parks, Social media usage is at an all-time high. that could mean a nightmare for democracy, Available at https://www.npr.org/2020/05/27/860369744/social-media-usage-is-at-an-all-time-high-that-could-mean-a-nightmare-for-democr?t=1656115935793 (2020).

29. E. Prapuolenis, Election 2020: Misinformation has never been more of a threat, Available at https://www.wsj.com/video/election-2020-misinformation-has-never-been-more-of-a-threat/CC10A237-BAFC-4F95-BD32-F6CA359C0C2C.html (2020).

30. A. Sebenius, Why disinformation is a major threat to the 2020 election, Available at https://www.washingtonpost.com/business/why-disinformation-is-a-major-threat-to-the-2020-election/2020/09/03/c0614588-ee17-11ea-bd08-1b10132b458f_story.html (2020).

31. A. C. Estes, R. Heilweil, Why mail-in voting conspiracy theories are so dangerous in 2020, Available at https://www.vox.com/recode/21454529/vote-by-mail-voter-fraud-trump-qanon-conspiracy-theory-2020-election (2020).

32. M. Graham, S. Rodriguez, Twitter and Facebook race to label a slew of posts making false election claims before all votes counted, Available at https://www.cnbc.com/2020/11/04/twitter-and-facebook-label-trump-posts-claiming-election-stolen.html (2020).

33. S. Alcorn, Behind the tweet that became the rallying cry for the insurrection, Available at https://revealnews.org/article/behind-the-tweet-that-became-the-rallying-cry-for-the-insurrection (2021).

34. P. Suciu, Social media proved crucial for Joe Biden – it allowed him to connect with young voters and avoid his infamous gaffes, Available at https://www.forbes.com/sites/petersuciu/2020/11/17/social-media-proved-crucial-for-joe-biden--it-allowed-him-to-connect-with-young-voters-and-avoid-his-infamous-gaffes (2020).

35. K. Roose, How Joe Biden's digital team tamed the MAGA internet, Available at https://www.nytimes.com/2020/12/06/technology/joe-biden-internet-election.html (2020).

36. J. C. Wong, One year inside Trump's monumental Facebook campaign, Available at https://www.theguardian.com/us-news/2020/jan/28/donald-trump-facebook-ad-campaign-2020-election (2020).

37. R. Mosquera, M. Odunowo, T. McNamara, X. Guo, R. Petrie, The economic effects of Facebook, *Experimental Economics* **23**, 575-602 (2019).

38. N. Asimovic, J. Nagler, R. Bonneau, J. A. Tucker, Testing the effects of Facebook usage in an ethnically polarized setting, *Proceedings of the National Academy of Sciences* **118** (2021).

39. A. M. Guess, *Advances in Experimental Political Science*, J. N. Druckman, D. P. Green, eds. (Cambridge University Press, 2021), p. 184–198.

40. C. A. Bail, *et al.*, Exposure to opposing views on social media can increase political polarization, *Proceedings of the National Academy of Sciences* **115**, 9216-9221 (2018).

677 41. R. Levy, Social media, news consumption, and polarization: Evidence from a field experi-
678 ment, *American Economic Review* **111**, 831-70 (2021).

679 42. M. Petrova, A. Sen, P. Yildirim, Social media and political contributions: The impact of
680 new technology on political competition, *Management Science* **67**, 2997–3021 (2021).

681 43. T. Rotesi, The impact of Twitter on political participation, *Working paper*, Bocconi Univer-
682 sity (2019).

683 44. K. Müller, C. Schwarz, From hashtag to hate crime: Twitter and antiminority sentiment,
684 *American Economic Journal: Applied Economics* **15**, 270–312 (2023).

685 45. T. Fujiwara, K. Müller, C. Schwarz, The effect of social media on elections: Evidence from
686 the United States, *Journal of the European Economic Association* (forthcoming).

687 46. N. Persily, J. A. Tucker, eds., *Social Media and democracy: The state of the field, prospects*
688 *for reform*, SSRC Anxieties of Democracy (Cambridge University Press, 2020).

689 47. J. A. Tucker, *et al.*, Social Media, Political Polarization, and Political Disinformation: A
690 Review of the Scientific Literature, *Political polarization, and political disinformation: A*
691 *review of the scientific literature* (2018).

692 48. J. Haidt, C. A. Bail, Social Media and Political Dysfunction: A Collaborative Review (on-
693 going). Unpublished manuscript, New York University.

694 49. S. González-Bailón, *et al.*, Asymmetric ideological segregation in exposure to political
695 news on facebook, *Science* **381**, 392–398 (2023).

696 50. A. M. Guess, *et al.*, How do social media feed algorithms affect attitudes and behavior in
697 an election campaign?, *Science* **381**, 398–404 (2023).

51. A. M. Guess, *et al.*, Reshares on social media amplify political news but do not detectably affect beliefs or opinions, *Science* **381**, 404–408 (2023).

52. B. Nyhan, *et al.*, Like-minded sources on facebook are prevalent but not polarizing, *Nature* pp. 1–8 (2023).

53. Pew Research Center, More than eight-in-ten Americans get news from digital devices, Available at https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/ (2021).

54. J. Allen, B. Howland, M. Mobius, D. Rothschild, D. J. Watts, Evaluating the fake news problem at the scale of the information ecosystem, *Science Advances* **6**, eaay3539 (2020).

55. H. Allcott, M. Gentzkow, Social Media and Fake News in the 2016 Election, *Journal of economic perspectives* **31**, 211–36 (2017).

56. A. M. Guess, B. Nyhan, J. Reifler, Exposure to untrustworthy websites in the 2016 US election, *Nature Human Behaviour* **4**, 472–480 (2020).

57. N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, D. Lazer, Fake news on Twitter during the 2016 US presidential election, *Science* **363**, 374–378 (2019).

58. E. Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think* (Penguin Books, 2011).

59. S. Rathje, J. J. Van Bavel, S. van der Linden, Out-group animosity drives engagement on social media, *PNAS* **118** (2021).

60. H. Hartig, 75% of Americans say it's likely that Russia or other governments will try to influence 2020 election, Available at https://www.pewresearch.org/fact-t

ank/2020/08/18/75-of-americans-say-its-likely-that-russia-or-other-governments-will-try-to-influence-2020-election (2020).

61. H. Hartig, The Trump Administration and the Media, Available at https://cpj.org/reports/2020/04/trump-media-attacks-credibility-leaks (2020).

62. M. Sullivan, Trump is leaving press freedom in tatters. Biden can take these bold steps to repair the damage., Available at https://www.washingtonpost.com/lifestyle/media/trump-biden-press-freedom/2020/12/04/251c669e-3583-11eb-a997-1f4c53d2a747_story.html (2020).

63. D. Kemp, E. Ekins, New poll: 75% don't trust social media to make fair content moderation decisions, 60% want more control over posts they see, Available at https://www.cato.org/blog/new-poll-75-dont-trust-social-media-make-fair-content-moderation-decisions-60-want-more (2020).

64. A. Akala, Don't censor! Stop the hoaxes! Facebook, Twitter face a catch-22, Available at https://www.npr.org/2020/10/16/924625825/dont-censor-stop-the-hoaxes-facebook-twitter-face-a-catch-22?t=1660562843875 (2020).

65. E. Ferrara, H. Chang, E. Chen, G. Muric, J. Patel, Characterizing social media manipulation in the 2020 US presidential election, *First Monday* **25** (2020).

66. M. Scott, R. Kern, The online world still can't quit the 'Big Lie', Available at https://www.politico.com/news/2022/01/06/social-media-donald-trump-jan-6-526562 (2022).

742 67. H. Homonoff, 2020 Political Ad Spending Exploded: Did It Work?, Available at https:
743 //www.forbes.com/sites/howardhomonoff/2020/12/08/2020-politi
744 cal-ad-spending-exploded-did-it-work/?sh=584a80d73ce0 (2020).

745 68. Y. Benjamini, A. M. Krieger, D. Yekutieli, Adaptive linear step-up procedures that control
746 the false discovery rate, *Biometrika* **93**, 491–507 (2006).

747 69. J. de Quidt, J. Haushofer, C. Roth, Measuring and Bounding Experimenter Demand, *Amer-*
748 *ican Economic review* **108** (2018).

749 70. D. Ghanem, S. Hirshleifer, K. Ortiz-Beccera, Testing Attrition Bias in Field Experiments,
750 *Journal of Human Resources* pp. 0920–11190R2 (2023).

751 71. H. Allcott, M. Gentzkow, L. Song, Digital addiction, *American Economic Review* **112**,
752 2424–2463 (2022).

753 72. S. DellaVigna, E. La Ferrara, *Handbook of Media Economics* (Elsevier, 2015), vol. 1, pp.
754 723–768.

755 73. D. A. Parry, *et al.*, A systematic review and meta-analysis of discrepancies between logged
756 and self-reported digital media use, *Nature Human Behaviour* **5**, 1535–1547 (2021).

757 74. D. Broockman, J. Kalla, The manifold effects of partisan media on viewers' beliefs and
758 attitudes: A field experiment with Fox News viewers (2022).

759 75. E. Oster, Unobservable Selection and Coefficient Stability: Theory and Evidence, *Journal*
760 *of Business & Economic Statistics* **37**, 187–204 (2019).

761 76. D. S. Lee, Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment
762 Effects, *The Review of Economic Studies* **76**, 1071–1102 (2009).