NBER WORKING PAPER SERIES

# THE WELFARE EFFECTS OF SOCIAL MEDIA

Hunt Allcott
Luca Braghieri
Sarah Eichmeyer
Matthew Gentzkow

Working Paper 25514
http://www.nber.org/papers/w25514

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2019

We thank Nancy Baym, Moira Burke, Annie Franco, Alex Leavitt, and Todd Rogers for helpful comments. We thank Kelly B. Wagman and Raj Bhargava for exceptional research assistance. We are grateful to the Sloan Foundation for grant funding. The study was approved by Institutional Review Boards at Stanford (eProtocol #45403) and NYU (IRB-FY2018-2139). This RCT was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0003409. Survey instruments are available from https://sites.google.com/site/allcott/research. Disclosures: Allcott is a paid employee of Microsoft Research. Gentzkow does paid consulting work for Amazon and is a member of the Toulouse Network for Information Technology, a research group funded by Microsoft. Braghieri and Eichmeyer have no relevant or material disclosures. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Welfare Effects of Social Media
Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow
NBER Working Paper No. 25514
January 2019
JEL No. D12,D90,I31,L86,O33

**ABSTRACT**

The rise of social media has provoked both optimism about potential societal benefits and concern about harms such as addiction, depression, and political polarization. We present a randomized evaluation of the welfare effects of Facebook, focusing on US users in the run-up to the 2018 midterm election. We measured the willingness-to-accept of 2,844 Facebook users to deactivate their Facebook accounts for four weeks, then randomly assigned a subset to actually do so in a way that we verified. Using a suite of outcomes from both surveys and direct measurement, we show that Facebook deactivation (i) reduced online activity, including other social media, while increasing offline activities such as watching TV alone and socializing with family and friends; (ii) reduced both factual news knowledge and political polarization; (iii) increased subjective well-being; and (iv) caused a large persistent reduction in Facebook use after the experiment. We use participants' pre-experiment and post-experiment Facebook valuations to quantify the extent to which factors such as projection bias might cause people to overvalue Facebook, finding that the magnitude of any such biases is likely minor relative to the large consumer surplus that Facebook generates.

Hunt Allcott
Department of Economics
New York University
19 W. 4th Street, 6th Floor
New York, NY 10012
and NBER
hunt.allcott@nyu.edu

Luca Braghieri
Economics Department
Stanford University
579 Serra Mall
Stanford, CA 94305
lucabrag@stanford.edu

Sarah Eichmeyer
Economics Department
Stanford University
579 Serra Mall
Stanford, CA 94305
saraeich@stanford.edu

Matthew Gentzkow
Department of Economics
Stanford University
579 Serra Mall Stanford,
CA 94305
and NBER
gentzkow@stanford.edu

An online appendix and survey instruments are available at
http://www.nber.org/data-appendix/w25514

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/3409

## 1 Introduction

Social media have had profound impacts on the modern world. Facebook, which remains by far the largest social media company, has 2.3 billion monthly active users worldwide (Facebook 2018). As of 2016, the average user was spending 50 minutes per day on Facebook and its sister platforms Instagram and Messenger (Facebook 2016). There may be no technology since television that has so dramatically reshaped the way people communicate, get information, and spend their time.

Speculation about social media's welfare impact has followed a familiar trajectory, with early optimism about potential benefits giving way to widespread concern about possible harms. At a basic level, social media dramatically reduce the cost of connecting, communicating, and sharing information with others. Given that interpersonal connections are among the most important drivers of happiness and well-being (Myers 2000; Reis, Collins, and Berscheid 2000; Argyle 2001; Chopik 2017), this could be expected to bring widespread improvements to individual welfare. Many have also pointed to wider social benefits, from facilitating protest and resistance in autocratic countries, to encouraging activism and political participation in established democracies (Howard et al. 2011; Kirkpatrick 2011).

More recent discussion has focused on an array of possible negative impacts. At the individual level, many have pointed to negative correlations between intensive social media use and both subjective well-being and mental health.[1] Adverse outcomes such as suicide and depression appear to have risen sharply over the same period that the use of smartphones and social media has expanded.[2] Alter (2018) and Newport (2019), along with other academics and prominent Silicon Valley executives in the "time well-spent" movement, argue that digital media devices and social media apps are harmful and addictive. At the broader social level, concern has focused particularly on a range of negative political externalities. Social media may create ideological "echo chambers" among like-minded friend groups, thereby increasing political polarization (Sunstein 2001, 2017; Settle 2018). Furthermore, social media are the primary channel through which fake news and other types of misinformation are spread online (Allcott and Gentzkow 2017), and there is concern that coordinated disinformation campaigns can affect elections in the US and abroad.

In this paper, we report on a large-scale randomized evaluation of the welfare impacts of Facebook, focusing on US users in the run-up to the November 2018 midterm elections. We recruited a sample of 2,844 users through Facebook display ads, and elicited their willingness-to-accept (WTA) to deactivate their Facebook accounts for a period of four weeks ending just after the election. We then randomly assigned the 58 percent of these subjects with WTA less than $102 to either a Treatment group that was paid to deactivate, or a Control group that was not. We verified compliance

[1] See, for example, Abeele et al. (2018), Burke and Kraut (2016), Ellison, Steinfield, and Lampe (2007), Frison and Eggermont (2015), Kross et al. (2013), Satici and Uysal (2015), Shakya and Christakis (2017), and Tandoc, Ferrucci, and Duffy (2015). See Appel, Gerlach, and Crusius (2016) and Baker and Algorta (2016) for reviews.

[2] See, for example, Twenge, Sherman, and Lyubomirsky (2016), Twenge and Park (2017), Twenge, Martin, and Campbell (2018), and Twenge et al. (2018).

with deactivation by regularly checking participants' public profile pages. We measured a suite of outcomes using text messages, surveys, emails, direct measurement of activity on Facebook and Twitter, and administrative records on voting and electoral contributions. Less than two percent of the sample failed to complete the endline survey, and the Treatment group's compliance with deactivation exceeded 90 percent.

Our study offers the largest-scale experimental evidence available to date on the way Facebook affects a range of individual and social welfare measures. We evaluate the extent to which time on Facebook substitutes for alternative online and offline activities, with particular attention to crowd out of news consumption and face-to-face social interactions. We study Facebook's broader political externalities via measures of news knowledge, awareness of misinformation, political engagement, and political polarization. We study the impact on individual utility via measures of subjective well-being, captured through both surveys and text messages. Finally, we analyze the extent to which behavioral forces like addiction and misprediction may cause sub-optimal consumption choices, by looking at how usage and valuation of Facebook change after the experiment.

Our first set of results focuses on substitution patterns. A key mechanism for effects on individual well-being would be if social media use crowds out face-to-face social interactions and thus deepens loneliness and depression (Twenge 2017). A key mechanism for political externalities would be if social media crowds out consumption of higher-quality news and information sources. We find evidence consistent with the first of these but not the second. Deactivating Facebook freed up 60 minutes per day for the average person in our Treatment group. The Treatment group actually spent less time on both non-Facebook social media and other online activities, while devoting more time to a range of offline activities such as watching television alone and spending time with friends and family. The Treatment group did not change its consumption of any other online or offline news sources and reported spending 15 percent less time consuming news.

Our second set of results focuses on political externalities, proxied by news knowledge, political engagement, and political polarization. Consistent with the reported reduction in news consumption, we find that Facebook deactivation significantly reduced news knowledge and attention to politics. The Treatment group was less likely to say they follow news about politics or the President, and less able to correctly answer factual questions about recent news events. Our overall index of news knowledge fell by 0.19 standard deviations. There is no detectable effect on political engagement, as measured by voter turnout in the midterm election and the likelihood of clicking on email links to support political causes. Deactivation significantly reduced polarization of views on policy issues and a measure of exposure to polarizing news. Deactivation did not statistically significantly reduce affective polarization (i.e. negative feelings about the other political party) or polarization in factual beliefs about current events, although the coefficient estimates also point in that direction. Our overall index of political polarization fell by 0.16 standard deviations. As a point of comparison, prior work has found that a different index of political polarization rose by

0.38 standard deviations between 1996 and 2018 (Boxell 2018).

Our third set of results looks at subjective well-being. Deactivation caused small but significant improvements in well-being, and in particular on self-reported happiness, life satisfaction, depression, and anxiety. Effects on subjective well-being as measured by responses to brief daily text messages are positive but not significant. Our overall index of subjective well-being improved by 0.09 standard deviations. As a point of comparison, this is about 25-40 percent of the effect of psychological interventions including self-help therapy, group training, and individual therapy, as reported in a meta-analysis by Bolier et al. (2013). These results are consistent with prior studies suggesting that Facebook may have adverse effects on mental health. However, we also show that the magnitudes of our causal effects are far smaller than those we would have estimated using the correlational approach of much prior literature. We find little evidence to support the hypothesis suggested by prior work that Facebook might be more beneficial for "active" users—for example, users who regularly comment on pictures and posts from friends and family instead of just scrolling through their news feeds.[3]

Our fourth set of results considers whether deactivation affected people's demand for Facebook after the study was over, as well as their opinions about Facebook's role in society. As the experiment ended, participants reported planning to use Facebook much less in the future. Several weeks later, the Treatment group's reported usage of the Facebook mobile app was about 12 minutes (23 percent) lower than in Control. The Treatment group was more likely to click on a post-experiment email providing information about tools to limit social media usage, and five percent of the Treatment group still had their accounts deactivated nine weeks after the experiment ended. Our overall index of post-experiment Facebook use is 0.61 standard deviations lower in Treatment than in Control. In response to open-answer questions several weeks after the experiment ended, the Treatment group was more likely to report that they were using Facebook less, had uninstalled the Facebook app from their phones, and were using the platform more judiciously. Reduced post-experiment use aligns with our finding that deactivation improved subjective well-being, and it is also consistent with the hypotheses that Facebook is habit forming in the sense of Becker and Murphy (1988) or that people learned that they enjoy life without Facebook more than they had anticipated.

Deactivation caused people to appreciate Facebook's both positive and negative impacts on their lives. Consistent with our results on news knowledge, the Treatment group was more likely to agree that Facebook helps people to follow the news. The great majority of the Treatment group agreed that deactivation was good for them, but they were also more likely to think that people would miss Facebook if they used it less. In free response questions, the Treatment group wrote more text about how Facebook has both positive and negative impacts on their lives. The opposing effects on these specific metrics cancel out, so our overall index of opinions about Facebook is unaffected.

[3] Correlation studies on active vs. passive Facebook use include Burke, Marlow, and Lento (2010), Burke, Kraut, and Marlow (2011), Burke and Kraut (2014), and Krasnova et al. (2013), and randomized experiments include Deters and Mehl (2012) and Verduyn et al. (2015).

Our work also speaks to an adjacent set of questions around how to measure the economic gains from free online services such as search and media.[4] In standard models with consumers who correctly optimize their allocation of time and money, researchers can approximate the consumer surplus from these services by measuring time use or monetary valuations, as in Brynjolfsson and Oh (2012), Brynjolfsson, Eggers, and Gannamaneni (2018), Corrigan et al. (2018), and others. But if users do not understand the ways in which social media could be addictive or make them unhappy, these standard approaches could overstate consumer surplus gains. Sagioglu and Greitemeyer (2014) provide suggestive evidence: while their participants predicted that spending 20 minutes on Facebook would make them feel better, it actually caused them to feel worse.

To quantify the possibility that a period of deactivation might help the Treatment group to understand ways in which their use had made them unhappy, we elicited WTA at three separate points, using incentive-compatible Becker-DeGroot-Marschak (1964, "BDM") mechanisms. First, on October 11th, we elicited willingness-to-accept to deactivate Facebook between October 12th and November 8th, which we loosely call "month 1." We immediately told participants the amount that they had been offered to deactivate ($102 for the Treatment group, $0 for Control), and thus whether they were expected to deactivate over that period. We then immediately elicited WTA to deactivate Facebook for the next four weeks *after* November 8th, which we call "month 2." When November 8th arrived, we then re-elicited WTA to deactivate in month 2. The Treatment group's change in valuation for month 2 reflects a time effect plus the unanticipated effect of spending time off of Facebook. The Control group's parallel valuation change reflects only a time effect. Thus, the difference between how Treatment vs. Control change their WTAs for deactivation in month 2 reflects projection bias, learning, and similar unanticipated experience effects, which we collectively call "misprediction."[5]

After weighting our sample to match the average US Facebook user on observables, the median and mean willingness-to-accept to deactivate Facebook for the initial four weeks were $100 and $180, respectively. These valuations are larger than most estimates in related work by Brynjolfsson, Eggers, and Gannamaneni (2018), Corrigan et al. (2018), Mosquera et al. (2018), and Sunstein (2019). Aggregated across an estimated 172 million US Facebook users, this could be interpreted to mean that Facebook generates several hundred billion dollars of consumer surplus per year in the US alone. Consistent with our other results that deactivation reduced demand for Facebook, deactivation caused month 2 WTA to drop by 13 percent, although this may be an upper bound on misprediction for reasons we discuss later. While such misprediction may be substantial in absolute

[4]See, for example, Brynjolfsson and Saunders (2009), Byrne, Fernald, and Reinsdorf (2016), Nakamura, Samuels, and Soloveichik (2016), Brynjolfsson, Rock, and Syverson (2018), and Syverson (2017).

[5]Our measurement of state dependence and misprediction connects us to the literature on habit formation and projection bias, including Acland and Levy (2015), Becker and Murphy (1988), Becker, Grossman, and Murphy (1991), Busse et al. (2015), Charness and Gneezy (2009), Conlin, O'Donoghue, and Vogelsang (2007), Fujiwara, Meng, and Vogl (2016), Gruber and Köszegi (2001), Hussam et al. (2016), Loewenstein, O'Donoghue, and Rabin (2003), and Simonsohn (2010).

terms, it would not reverse the conclusion that Facebook generates enormous flows of consumer surplus.

Our results should be interpreted with caution, for several reasons. First, effects could differ with the duration or scale of deactivation. A longer period without Facebook might have less impact on news knowledge as people find alternative news sources, and either more or less impact on subjective well-being. Furthermore, a larger-scale experiment in which a greater share of the population deactivated could have a different impact due to network effects and equilibrium adjustments. Second, our sample is not fully representative. Our participants are relatively young, well-educated, and left-leaning compared to the average Facebook user, and we included only people who reported using Facebook more than 15 minutes per day. In addition, although we went as far as possible to avoid telegraphing the experimental design and research questions, deactivation could have different effects on the average Facebook user than on the type of person who was willing to participate in our experiment. Third, many of our outcome variables are self-reported, adding scope for both measurement error and experimenter demand effects. This latter concern is mitigated somewhat by the fact that the non-self-reported outcomes we measure (e.g., post-experiment Facebook use) paint a similar picture to the survey responses.

The causal impacts of social media have been of great interest to researchers in economics, psychology, and other fields. We are aware of 12 existing randomized impact evaluations of Facebook.[6] The most closely related is the important paper by Mosquera et al. (2018), which was made public the month before ours. They also use Facebook deactivation to study news knowledge and well being, finding results broadly consistent with those reported here. Appendix Table A1 details these experiments in comparison to ours. Our deactivation period is substantially longer and our sample size an order of magnitude larger than most prior work, including Mosquera et al. (2018). We measure impacts on a relatively comprehensive range of outcomes, and we are the only one of these randomized trials to have submitted a pre-analysis plan. Given the effect sizes and residual variance in our sample, we would have been unlikely to have sufficient power to detect any effects if limited to the sample sizes in previous experiments.

Sections 2 and 3 detail the experimental design and empirical strategy. Section 4 presents the impact evaluation, and Section 5 presents measurements of the consumer surplus generated by Facebook.

[6] These studies sit within a broader media effects literature that uses experimental and quasi-experimental methods to quantify the effects of media technologies such as television, media providers such as Fox News, and content such as political advertising (Bartels 1993; Besley and Burgess 2001; DellaVigna and Kaplan 2007; Enikolopov, Petrova, and Zhuravskaya 2011; Gentzkow 2006; Gerber and Green 2000; Gerber et al. 2011; Gerber, Karlan, and Bergan 2009; Huber and Arceneaux 2007; Martin and Yurukoglu 2017; Olken 2009; and Spenkuch and Toniatti 2016); for reviews, see DellaVigna and Gentzkow (2010), Napoli (2014), Strömberg (2015), Enikolopov and Petrova (2015), and DellaVigna and La Ferrara (2015).

# 2 Experimental Design

## 2.1 Experiment Overview

Figure 1 summarizes our experimental design and timeline. We timed the experiment so that the main period of Facebook deactivation would end shortly after the 2018 US midterm elections, which took place on November 6th. The experiment has eight parts: recruitment, pre-screen, baseline survey, midline survey, endline survey, post-endline survey, post-endline emails, and daily text messages. Table 1 presents the sample sizes at each part of the experiment.

Between September 24th and October 3rd, we recruited participants using Facebook ads. Our ad said, "Participate in an online research study about internet browsing and earn an easy $30 in electronic gift cards." Appendix Figure A1 presents the ad. To minimize sample selection bias, the ad did not hint at our research questions or suggest that the study was related to social media or Facebook deactivation. We targeted the ads by demographic cells in an attempt to gather an initial sample that was approximately representative of Facebook users on gender, age, college completion, and political ideology. 1,690,076 unique users were shown the ad, of whom 30,064 clicked on it. This 1.8 percent click-through rate is about twice the average click-through rate on Facebook ads across all industries (Irvine 2018).

Clicking on the ad took the participant to a brief pre-screen survey, which included several background demographic questions and the consent form. The pre-screen, baseline, midline, and endline surveys were hosted on the site stanforduniversity.qualtrics.com. 14,324 people passed the pre-screen, by reporting being a US resident born between the years 1900 and 2000 who uses Facebook more than 15 minutes and no more than 600 minutes per day. Of those people, 5,974 consented to participate in the study.

After completing the consent form, participants began the baseline survey. The baseline recorded email addresses, additional demographics, and a range of outcome variables. We also asked for each participant's name, zip code, Twitter handle, and phone number ("in order for us to send you text messages during the study"), as well as the URL of their Facebook profile page (which we would use "solely to observe whether your Facebook account is active").

To minimize selective attrition, we asked all participants regardless of subsequent treatment status to deactivate their Facebook accounts for 24 hours following the midline and endline surveys. In the baseline, we informed people that "As part of this study, we will ask you to deactivate your Facebook account twice for a period of 24 hours. You will keep your access to Facebook Messenger. If you deactivate, you can choose to come back whenever you want with your content and friends network unchanged," and asked, "Are you willing to deactivate your Facebook account twice for 24 hours (once after Survey 2 and once after Survey 3)?"

In all, 3,234 people finished the baseline survey and responded that they were willing to deactivate. Of those, 183 were dropped from the experimental sample because of invalid data (for example, invalid Facebook profile URLs) or low-quality baseline responses (for example, discrepancies between average daily Facebook usage reported in the pre-screen vs. baseline survey, completing the survey in less than ten minutes, no text in short-answer boxes, and other patterns suggesting careless responses). The remaining 3,051 participants had valid baseline data, were included in our stratified randomization, and were invited to take the midline survey.

On October 11th, we sent email invitations to the midline survey. In that survey, we first asked people to deactivate their Facebook account for 24 hours, and guided them through the process. Second, we used a Becker-DeGroot-Marschak (BDM) mechanism to elicit willingness-to-accept (WTA) to stay deactivated for four weeks rather than 24 hours.[7] We then revealed the BDM price offer. Participants whose WTA was strictly less than the price draw were informed that they should deactivate for the full four weeks after midline. Third, we reminded people that we would again ask them to deactivate for 24 hours after the endline survey, and used a second BDM mechanism to elicit WTA to stay deactivated for the four weeks after endline instead of just 24 hours. We informed people that "We will check continuously whether your account is deactivated for the entire [24 hours/4 weeks] in which it is supposed to be by pinging the URL associated with your profile." For expositional purposes, we will loosely refer to the four weeks after midline as "month 1," and the four weeks after endline as "month 2."

On November 8th, two days after the midterm election, we sent an email invitation to the endline survey. The endline survey first measured the same outcome variables as the baseline survey. All questions were identical, with the exception of cases discussed in Section 2.3 below, such as using updated news knowledge questions and rephrasing questions about the midterm election to be in the past tense. We then asked all participants to again deactivate their Facebook accounts for the next 24 hours, and again elicited WTA to stay deactivated for the next four weeks (i.e., month 2) instead of the next 24 hours. Participants were told, "With a 50% chance we will require you to abide by the decision you made 4 weeks ago; with 50% chance we will ignore the decision you made 4 weeks ago and we will require you to abide by the decision you make today."

We gathered data from two post-endline emails. On November 20th, we sent an email with links to information on ways to limit smartphone social media use, and on November 25th, we sent an email with links to donate to, volunteer for, or sign petitions related to political causes. Clicks on these emails provide additional non-self-reported measures of interest in reducing social media use and political engagement. Appendix Figures A2 and A3 present the two emails.

[7] The survey explained, "The computer has randomly generated an amount of money to offer you to deactivate your Facebook account for the next 4 weeks. Before we tell you what the offer is, we will ask you the smallest offer you would be willing to accept. If the offer the computer generated is above the amount you give, we will ask you to deactivate for 4 weeks and pay you the offered amount if you do. If the offer is below that amount, we will not ask you to deactivate." We then asked several comprehension questions to make sure that participants understood the mechanism. We did not tell participants the distribution or support of the offer prices, both because we did not want to artificially truncate the distribution of elicited WTA and because prior studies have found that providing information on the bounds of the offer price distribution can affect BDM valuations (Bohm, Lindén, and Sonnegård 1997; Mazar, Koszegi, and Ariely 2014).

On December 3rd, we invited participants to a short post-endline survey in which we asked how many minutes per day they had used the Facebook app on their smartphones in the past seven days. We asked participants with iPhones to report the Facebook app time reported by their phone's Settings app, and we asked other participants to estimate. We also asked several open-answer questions, such as "How has the way you use Facebook changed, if at all, since participating in this study?"

For the approximately six weeks between baseline and endline, we sent daily text message surveys to measure several aspects of subjective well-being in real time rather than retrospectively. We rotated three types of questions, measuring happiness, the primary emotion felt over the past ten minutes, and loneliness. Appendix Figure A4 presents the three questions.

We verified deactivation by checking each participant's Facebook profile page URL regularly at random times. While a user can limit how much content other people can see in their profiles, they cannot hide their public profile page, and the public profile URL returns a valid response if and only if their account is active.[8] This is thus our measure of deactivation. For all participants, we verified deactivation approximately once per day for the seven days before midline and all days between endline and late January 2019. Between midline and endline, we verified deactivation approximately four times per day for the participants who had been randomly assigned to deactivate (i.e., the Treatment group) and once every four days for participants who had not been assigned to deactivate. During the post-midline and post-endline 24-hour deactivation periods, we generally verified deactivation within about six hours of when each participant completed the survey. If participants were not deactivated when they were supposed to be, our program immediately sent an automated email informing them that they should again deactivate as soon as possible, along with a survey asking them to explain why they were not deactivated. We discuss the way we handle imperfect compliance in our empirical analysis in Section 3 below.

All participants received $5 per completed survey, paid via gift card immediately upon completion. All participants were told that they would receive a $15 "completion payment" if they completed all surveys, responded to 75 percent of text messages, kept their accounts deactivated for the 24 hours after midline and endline, and, if the deactivation offer price was above their reported WTA, kept their accounts deactivated for the full period between midline and endline. The latter requirement (making the completion payment contingent on complying with the BDM's deactivation assignment) makes it a strictly dominant (instead of weakly dominant) strategy to truthfully report

[8]By default, Facebook profile URLs end in a unique number, which is the numeric ID for that person in the Facebook system. Users can update their default URL to be something customized, and they can change their customized URL as often as they want. In the baseline survey, participants reported their profile URLs, which could have been either the default or customized version. Shortly after the baseline survey, we checked if each participant's Facebook profile URL was valid by pinging it and looking in the page source for the string containing the person's numeric ID. If the numeric ID existed, we knew that the URL was valid. After that point, we used participants' numeric IDs to construct their default numeric URLs, which allowed us to correctly measure deactivation even if they changed their customized URL.

valuations in the BDM.[9] These payments were in addition to the \$102 that treatment participants received in exchange for deactivation.

## 2.2 Randomization

We use the BDM mechanism described above to randomly assign participants to Facebook deactivation. Figure 1 illustrates the randomization. Participants with valid baseline data were randomized into three groups that determined the BDM offer price $p$ for deactivation in month 1 (i.e., the weeks between midline and endline): $p = \$102$ (approximately 35 percent of the sample), $p = \$0$ (approximately 65 percent), and $p$ drawn from a uniform distribution on $[\$0, \$170]$ (approximately 0.2 percent).[10] We balanced the $p = \$102$ and $p = \$0$ group assignments within 48 strata defined by age, average daily Facebook use, heavy vs. light news use (those who get news from Facebook fairly often or very often vs. never, hardly ever, or sometimes), active vs. passive Facebook use, and Democrat, Republican, or independent party affiliation.

The effects of Facebook deactivation in month 1 are identified in the sample of participants who were allocated to $p = \$102$ or $p = \$0$ and were willing to accept less than \$102 to deactivate in month 1. We call this the "impact evaluation sample." Within the impact evaluation sample, we call $p = \$102$ the "Treatment" group, and $p = \$0$ the "Control" group.

For deactivation in month 2 (i.e., the four weeks after endline), 0.2 percent of participants were randomly selected to a BDM offer price drawn randomly from $p' \in [0, 170]$, while the remaining 99.8 percent received offer $p' = 0$. We balanced this month 2 offer price $p'$ between the month 1 offer price groups, so two participants who were offered $p = \$102$ and four participants who were offered $p = \$0$ were assigned to positive month 2 offers $p' \in [0, 170]$.

This approach allows us to maintain incentive compatibility in the BDM mechanism, have balance between Treatment and Control groups, and use a simple regression to estimate treatment effects of post-midline deactivation.

## 2.3 Outcome Variables

For the impact evaluation, we consider the outcome variables in the nine families described below. Appendix B.1 presents survey question text and descriptive statistics for each outcome variable and moderator, grouped by family. We also construct indices that combine the outcome variables within each family, weighting by the inverse of the covariance between variables at endline, as described in Anderson (2008). In constructing these indices, we orient the variables so that more

[9] As discussed above, we did not inform participants of the BDM offer price distribution. Thus, more precisely, truthfully reporting valuations is a strictly dominant strategy only within the support of the offer price distribution that participants expected us to use.

[10] \$170 was chosen because it was the maximum that we could pay participants without requiring tax-related paperwork.

positive values have the same meaning—for example, more positive means "more polarized" in all cases. Outcomes to be multiplied by -1 are followed by "$\times$ (-1)" in Appendix B.1.

**Substitute time uses**

At baseline and endline, we asked participants how many minutes per day they spent on Facebook on the average day in the past four weeks. At baseline, we also asked participants to report how much of their free time on the average day in the past four weeks they spent on various activities, ranging from using social media apps other than Facebook to spending time with friends and family in person. At endline, we asked how much time they spent on the same activities, "relative to what is typical for you." We phrased the questions in this way in order to more precisely detect changes in self-reported time use caused by the deactivation.

**Social interaction**

We have three measures of social interaction. The *friends met in person* variable is the natural log of one plus the number of friends seen in person in the last week, as measured by a survey question that asked participants to "list the first names of as many friends you met in person last week that you can think of in 1 minute." *Offline activities* is the number of offline activities (such as going out to dinner, spending time with your kids, etc.) that the person did at least once last week. *Diverse interactions* is an indicator for whether the respondent interacted with someone who voted the opposite way in the last presidential election plus an indicator for whether the respondent interacted with someone from another country in the last week.

**Substitute news sources**

At baseline, we asked participants how often they got news from different sources over the past four weeks, including Facebook, cable TV, print, and radio news, borrowing a standard survey question from the Pew Research Center (2018a). At endline, we again asked how often they got news from those same sources, "relative to what is typical for you." For the participants who reported having a Twitter handle, we gathered data on number of tweets in the four weeks before baseline began and in the four weeks between midline and endline. This allows a non-self-reported measure of one kind of potential substitution away from Facebook.[11]

**News knowledge**

In order to detect broad changes in news exposure, we asked participants how closely they followed politics, how closely they followed news about President Trump, and how many minutes per day

[11] In our pre-analysis plan, we grouped this *number of tweets* variables in the substitute news sources family, but one might also think of it as a "substitute time use" because Twitter is not only used to read news.

they spent watching, reading, or listening to the news (including on social media) over the past four weeks.

In order to measure specific news knowledge, we included a 15-question news knowledge quiz. For each question, we gave a statement from the news in the past four weeks and asked participants to indicate if they thought the statement was true or false, or whether they were unsure. The order of the 15 statements was randomized. Seven of the statements were from news stories covered in the past four weeks in six news websites: New York Times, Wall Street Journal, Fox News, CNN, MSNBC, and US News & World Report, such as "The Trump administration set the maximum number of refugees that can enter the country in 2019 to 30,000." Three of the headlines were false modifications of articles from those same six news websites, such as "President Trump spoke at the funeral of former Arizona Senator John McCain, honoring the late McCain's wish." (In reality, it had been reported that President Trump was not invited to McCain's funeral.) The *news knowledge* variable is the count of true statements rated as true plus the count of false statements rated as false, plus one-half for every statement about which the respondent was "unsure." The final five statements were from fake news stories—rated false by third-party fact-checkers snopes.com and factcheck.org—that circulated heavily within a four-week period before the survey. The *fake news knowledge* variable is the count of fake statements correctly rated as "false" plus one-half for every statement about which the respondent was unsure. Appendix B presents the full news knowledge quizzes from both baseline and endline.

### Political engagement

We have two measures of political engagement. First, we measure whether participants voted in the 2018 midterm election using voting self-reports from the endline survey. Second, we measure whether participants clicked on any of the links in the post-endline politics email.

### Political polarization

There are a variety of ways to measure political polarization (see, for example, Gentzkow 2016), and we use both standard and novel measures. First, we included standard "feeling thermometer" questions capturing how "warm or cold" participants felt toward the Demoratic and Republican Parties and President Trump over the past four weeks. The *party affective polarization* variable is the respondent's thermometer warmth toward her own party minus her warmth toward the other party. Second, the *Trump affective polarization* variable is the thermometer warmth toward President Trump for Republicans, and minus one times the thermometer warmth toward President Trump for Democrats. For these and all other polarization variables, we include independents who lean toward a party, and we drop independents who do not lean toward either party. Third, we asked respondents to list recent news events that made them angry at the Republican or Democratic Party.

*Party anger* is the natural log of one plus the length (in characters of text) of her response about the other party minus the natural log of one plus the length of her response about her own party. Fourth, *other party understanding* is the number of times the respondent reported seeing news that made her better understand the point of view of the other political party minus the number of times she saw news that helped her understand her own party. Fifth, we asked nine opinions about current political issues, such as "To what extent do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States?" The *issue polarization* variable reflects the extent to which the respondent's issue opinions align with the average opinion in her own party instead of the other party. Sixth, *belief polarization* reflects the extent to which the respondent's beliefs about current news events (from the news knowledge quiz described above) align with the average belief in her own party instead of the other party.[12] Finally, *vote polarization* measures the strength of preferences for the congressional candidate of the respondent's party in the midterm election.[13]

**Subjective well-being**

There is a vast literature on measuring subjective well-being (see, for example, Kahneman et al. 2006), and we use standard measures from the literature. We modified existing scales in two ways. First, we asked questions in reference to the past four weeks, so as to increase our ability to detect changes as a result of Facebook deactivation. Second, in some cases we chose a subset of questions from standard multi-question scales in order to focus on areas of subjective well-being that might be most affected by Facebook.

The *happiness* variable is the average response to two questions from the Subjective Happiness Scale (Lyubomirsky and Lepper 1999), asking how happy participants were over the past four weeks and how happy they were compared to their peers. *Life satisfaction* is the sum of responses to three questions from the Satisfaction with Life Scale (Diener et al. 1985), such as the level of agreement with the statement, "During the past 4 weeks, I was satisfied with my life." *Loneliness*

[12] Specifically, for each issue or belief question $q$, we normalize responses by the standard deviation in the Control group, determine Democrats' and Republicans' average responses $\mu_q^D$ and $\mu_q^R$, re-center so that $\mu_q^D + \mu_q^R = 0$, and re-sign so that $\mu^R > 0$. Define $\tilde{y}_{iq}$ as individual $i$'s normalized, re-centered, and re-signed response to question $q$, multiplied by -1 if $i$ is a Democrat. $\tilde{y}_{iq}$ thus reflects the strength of individual $i$'s agreement with the average view of her party instead of the other party. For *issue polarization*, further define $\sigma_q$ as the Control group within-person standard deviation of $\tilde{y}_{iq}$ for question $q$. This measures how much people's views change between baseline and endline, and allows us to place higher weight on issues about which views are malleable over the deactivation period. For belief polarization, let $\sigma_q = 1$. The issue and belief polarization measures are $Y_i = \sum_q \tilde{y}_{iq}\sigma_q$.

[13] Specifically, we asked "In the recent midterm elections, did you vote for the Republican Party's or for the Democratic Party's candidate for Congress in your district? (If you did not vote, please tell us whom you would have voted for.)" We code vote polarization as 0 for "other/don't know." For people who responded that they had (or would have) voted for the Republican or Democratic candidate, we then asked, "How convinced were you about whether to vote for the Republican candidate or the Democratic candidate?" In these cases, we code vote polarization on a scale from -1 (very convinced to vote for the Democratic candidate) to +1 (very convinced to vote for the Republican candidate), and then multiply by -1 for Democrats.

is the Three-Item Loneliness Scale (Hughes et al. 2004). Finally, *depressed*, *anxious*, *absorbed*, and *bored* reflect how much of the time during the past four weeks respondents felt each emotion, using questions from the European Social Survey well-being module (Huppert et al. 2009).

The daily text messages allowed us to measure the aspects of subjective well-being that are most important to record in the moment instead of retrospectively. This approach builds on the Experience Sampling Method of Csikszentmihalyi and Larson (2014) and Stone and Shiffman (1994). The variable *SMS happiness* is the answer to the question, "Overall, how happy do you feel right now on a scale from 1 (not at all happy) to 10 (completely happy)?" The variable *SMS positive emotion* is an indicator variable for whether the participant reports a positive emotion when asked, "What best describes how you felt over the last ten minutes?", with possible responses such as "angry," "worried," "loving/tender," etc. Finally, *SMS not lonely* uses the answer to the question, "How lonely are you feeling right now on a scale from 1 (not at all lonely) to 10 (very lonely)?"

**Post-experiment Facebook use**

We have four measures of planned and actual post-experiment Facebook use. First, *planned post-study use change* is the extent to which participants plan to use Facebook more or less than they had before they started the study. (This was included only in the endline survey.) Second, *clicked time limit email* is an indicator for whether the respondent clicked any of the links in the post-endline social media time limit email. Third, *speed of reactivation* is minus one times the natural log of one plus the number of days that the participant's account remained deactivated between the post-endline 24-hour deactivation period and our most recent measurement on December 17th. Fourth, *Facebook mobile app use* is the natural log of one plus the number of minutes per day that the participant reported using Facebook on their phone in the post-endline survey.

**Opinions about Facebook**

We asked eight questions eliciting people's opinions about Facebook, such as "To what extent do you think Facebook is good or bad for society?" and "To what extent do you think Facebook makes people more or less politically polarized?" Each of these eight responses was on a ten-point scale. In the endline survey only, we also asked *deactivation bad*: "As part of this study, you were asked to deactivate your Facebook account for [24 hours/4 weeks]. To what extent do you think that deactivating your account was good or bad for you?" Finally, we also included two open answer text boxes in which we asked people to write out the most important positive and negative impacts that Facebook has on their lives. The *positive impacts* and *negative impacts* variables are the natural log of one plus the count of characters in the respective text box.

**Secondary outcomes**

We also consider the following two outcomes, which we labeled as "secondary" in our pre-analysis plan.

First, we consider the standard generic ballot question: "If the elections for US Congress were being held today, would you vote for the Republican Party's candidate or the Democratic Party's candidate for Congress in your district?" To increase precision, we then asked, "How convinced are you about whether to vote for the Republican or Democratic candidate?" At endline, we asked these questions in past tense, about whom the respondent did vote for in the 2018 midterm (or whom the respondent would have voted for had she voted, to avoid potentially selective non-response). The *voted Republican* variable is the strength of preferences for the Republican candidate. We labeled this outcome as secondary because we expected the estimates to be too imprecise to be of interest.

Second, we plan to gather contributions to political campaigns from the Federal Election Commission (FEC) database, both in the four years before baseline and during the midline-endline period. We labeled this as secondary because very few Americans contribute to political campaigns, and we did not expect to be able to detect effects of four weeks of deactivation. This outcome is not included in our analysis because the FEC database has not yet been updated for the relevant time period.

# 3 Empirical Strategy

## 3.1 Sample Demographics, Response Rates, and Compliance with Deactivation

Table 1 shows sample sizes at each step of our experiment, from the 1.7 million Facebook users who were shown our ads, to the 1,661 subjects in the impact evaluation sample. Table 2 quantifies the representativeness of our sample on observables, by comparing the demographics of our impact evaluation sample to our estimate of the average demographics of adult Facebook users and to the US adult population. Comparing column 1 to columns 2 and 3, we see that our sample is relatively high-income, well-educated, female, young, and Democratic, and uses Facebook relatively heavily.[14] Appendix Table A10 shows that Treatment and Control are balanced on observables.

Table 3 documents very high response rates to the endline and post-endline surveys and subjective well-being text messages. Of the 580 people in the Treatment group, only seven failed to complete the endline survey. Of the 1,081 people in the Control Group, only 17 failed to complete endline. The average participant responded to 92 percent of daily text messages, well above the

[14]In Appendix Figures A17, A18, A19, and A20, we find that the two demographic variables that we pre-specified as moderators, age and political party, do not appear to systematically moderate treatment effects. Furthermore, Figure 9 provides no systematic evidence that the effects vary for people who use Facebook more vs. less heavily before baseline. This suggests that re-weighting the sample for representativeness on these observables would not substantively change the estimated effects, although it would increase the standard errors.

75 percent required in order to receive the completion payment.[15] Treatment and Control have statistically equal response rates to the endline survey and subjective well-being text messages. A marginally significantly larger share of the Treatment group responded to the post-endline survey; this is less worrisome because Facebook mobile app use is the only variable from that survey for which we calculate treatment effects, and we show in Appendix Table A9 that using Lee (2009) bounds to account for attrition does not change the conclusions. Finally, Table 3 also reports the high level of compliance with our deactivation treatment: Treatment group participants were deactivated on 90 percent of checks between 24 hours after midline and the day they completed the endline survey, against two percent for Control.

As described above, if Treatment group members were found to have active accounts, we sent an email informing them of this and asking them to promptly deactivate, along with a survey asking why they were not deactivated. From these surveys, along with email interactions and formal qualitative interviews following our summer 2018 pilot study, we conclude that most Treatment group members who did reactivate fall into one of two groups. The first group consists of a small number of users who changed their mind about participating in the experiment and reactivated intentionally. The second group consists of users who briefly reactivated by accident, for example because they logged in to another app or online service using their Facebook account credentials.[16]

Appendix Figure A27 shows the cumulative distribution of the share of time deactivated, and Appendix Figure A28 shows the distribution of reasons for deactivation among those for whom this share was less than one. Together, these figures suggest that the small group of intentional reactivators accounts for the vast majority of Treatment group non-compliance. Given this, combined with the fact that only a small share of the Control group also remains deactivated (Figure 7), we estimate treatment-on-the-treated effects using instrumental variables.

### 3.2 Pre-Analysis Plan

We submitted our pre-analysis plan on October 12, immediately after the midline survey was completed. We submitted a slightly updated pre-analysis-plan on November 7, the day before endline, with only one substantive change: on the basis of data on reasons for non-compliance described above, we specified that our primary specifications would use IV estimates instead of intent-to-treat estimates. The pre-analysis plan specified three things. First, it specified the outcome variables and families of outcome variables as described above, including which specific variables are included in the index for each family and which outcomes are "secondary." Versions of Figures 2, 3, 5, 6, 7,

[15] Appendix Figure A26 shows the text message response rate by day (response rates declined slightly over the course of the experiment) and shows that Treatment and Control response rates are statistically balanced in all days of the deactivation period.

[16] Users who use their Facebook credentials to log in to another app such as Spotify can continue to use the app provided they were logged in before deactivation. If they logged in during the deactivation period, however, this would re-activate their Facebook account. We gave subjects a prominent warning about this in the deactivation instructions.

and 11 appear as figure shells in the pre-analysis plan, although we changed the order in which we present the families of outcome variables for expositional purposes. Second, the pre-analysis plan specified the moderators we use when testing for heterogeneous treatment effects, including which moderators are "secondary." Third, it specified the two regression specifications and the estimation sample as described below.

### 3.3 Empirical Strategy

To estimate the local average treatment effect (LATE) of Facebook deactivation, define $Y_i$ as some outcome measured at endline, and $\boldsymbol{Y}_i^b$ as a vector including the baseline value of the outcome and the baseline value of the index that includes the outcome.[17] Define $D_i$ as the percent of deactivation checks between (and including) October 12th (the day after midline) and November 7th (the day before endline) that person $i$ is observed to be deactivated. Define $T_i \in \{1, 0\}$ as a Treatment group indicator, and $\nu_s$ as the vector of the 48 stratum dummies. We estimate local average treatment effects of deactivation using the following regression:

$$Y_i = \tau D_i + \rho \boldsymbol{Y}_i^b + \nu_s + \varepsilon_i, \tag{1}$$

instrumenting for $D_i$ with $T_i$. In Equation (1), $\tau$ measures the local average treatment effect of deactivation for people induced to deactivate by the experiment.

Facebook deactivation might have a larger impact for people who use Facebook more. Define $H_i$ as person $i$'s average daily hours of Facebook use reported at baseline, winsorized at 120 minutes. We can also estimate the local average treatment effect of deactivation *per hour of daily Facebook use avoided* using the following regression:

$$Y_i = \tau D_i H_i + \beta H_i + \rho \boldsymbol{Y}_i^b + \nu_s + \varepsilon_i, \tag{2}$$

analogously instrumenting for $D_i H_i$ with $T_i H_i$.

If effects of deactivation are indeed linear in avoided hours of Facebook use, then Equation (2) could provide more statistical power than Equation (1). On the other hand, if effects are closer to constant in baseline usage and/or $H_i$ is measured with error, then Equation (1) will offer more power. In our pre-analysis plan, we specified that we would make either Equation (1) or Equation (2) our primary specification, depending on which delivered more power. In reality, the results are very similar. Therefore, we focus on Equation (1) because it is simpler. Appendix D presents results using Equation (2).

The base sample for all regressions is the "impact evaluation sample"—again, participants who

[17] $\boldsymbol{Y}_i^b$ excludes the baseline value of the outcome for outcomes such as clicks on post-endline emails that do not have a baseline value. $\boldsymbol{Y}_i^b$ excludes the baseline index when $Y_i$ is not included in an index. When $Y_i$ is an index, $\boldsymbol{Y}_i^b$ is simply the baseline value of the index.

were willing to accept less than \$102 to deactivate in month 1 (the four weeks after midline) and were offered $p = \$102$ or $p = \$0$ to do so. For the political polarization outcomes, the sample includes only Democrats and Republicans, as well as independents who lean toward one party or the other. Sample sizes sometimes differ across outcomes due to missing data: for example, the post-endline survey has slightly higher non-response than the endline survey, and many participants do not have Twitter accounts.

We use robust standard errors in all regressions.

# 4 Impact Evaluation

This section presents treatment effects of Facebook deactivation. The following subsections present estimates for four groups of outcomes: substitution, news and political outcomes, subjective well-being, and post-experiment Facebook use and opinions. We then present heterogeneous treatment effects.

In the body of the paper, we present figures with local average treatment effects and 95 percent confidence intervals from estimates of Equation (1), with outcome variables $Y_i$ normalized so that the Control group standard deviation equals one. Appendix Tables A6 and A7 provide numerical regression results for all individual outcome variables in both normalized (standard deviation) units, as in the figures, and un-normalized (original) units. Appendix Table A8 provides numerical regression results for all nine summary indices. These appendix tables also provide unadjusted p-values and "sharpened" False Discovery Rate (FDR)-adjusted p-values following the procedure of Benjamini, Krieger, and Yekutieli (2006), as outlined by Anderson (2008). The unadjusted p-values are appropriate for readers with a priori interest in one specific outcome. The FDR-adjusted p-values for the individual outcomes limit the expected proportion of false rejections of null hypotheses across all individual outcomes reported in the paper, while the FDR-adjusted p-values for the indices limit the expected proportion of false rejections of null hypotheses across the nine indices. The sharpened FDR-adjusted p-values are less conservative than the unadjusted p-values for p-values greater than about 0.15, and more conservative for unadjusted p-values less than that.

## 4.1 Substitutes for Facebook

Figure 2 presents treatment effects on substitutes for Facebook: substitute time uses, social interactions, and substitute news sources. Substitution is of interest for two reasons. First, our treatment entails deactivating Facebook *and* also re-allocating that time to other activities. Understanding that re-allocation is thus crucial for conceptually understanding the "treatment." Second, this substitution helps to understand mechanisms for key effects. One central mechanism through which Facebook might affect psychological well-being is by crowding out face-to-face interactions. However, it's also possible that when people deactivate, they primarily devote their newly available time to other solitary pursuits. Furthermore, a central mechanism for possible political externalities is that social media crowds out consumption of higher-quality news. However, it's also possible that when people deactivate, they simply get less news overall instead of substituting to other news sources.

The top group of outcomes in Figure 2 measures self-reported time use. Facebook usage was reported in minutes. For all other activities, the endline survey asked respondents how much time they spent on the activity in the last four weeks relative to what is typical for them, on a five-point scale from "A lot less" to "A lot more." For all time use outcomes, "Same" is the average answer in the Control group.

The first row confirms that the treatment indeed reduced Facebook use as intended. At endline, the Control group reported that they had used Facebook for an average of 59.53 minutes per day over the past four weeks, and the local average treatment effect of deactivation is 59.58 minutes per day.[18] As shown on Figure 2, this corresponds to a reduction of 1.59 standard deviations.

We find that Facebook deactivation *reduced* time devoted to other online activities. Time using non-Facebook social media falls by a quarter point on our five-point scale (0.27 SD), and time on non-social online activities falls by 0.12 points (0.14 SD). Thus, Facebook appears to be a complement rather than a substitute for other online activities. This makes sense to the extent that deactivating Facebook makes people less likely to be using their phones or computers in the first place, and less likely to follow Facebook links that direct to non-Facebook sites (e.g., a news website or Twitter post). It may also reflect the fact that many people use their Facebook credentials to log in to other apps and services such as Spotify and Tinder.[19]

Rows 4-7 of Figure 2 suggest that the 60 minutes freed up by not using Facebook, as well as the additional minutes from reductions in other online activities, were allocated to both solitary and social activities offline. Solitary television watching increases by 0.17 points on our scale (0.17 SD), other solitary offline activities increase by 0.23 points (0.25 SD), and time devoted to spending time with friends and family increases by 0.14 points (0.16 SD). The substitute time uses index, which does not include Facebook minutes, shows an increase in overall non-Facebook activities. All of the online and offline time use effects are highly significant with and without adjustment for multiple hypothesis testing.

The middle group of outcomes in Figure 2 contains measures of social interaction. Deactivation increased the count of offline activities that people reported doing at least once last week by about 0.18 (0.12 SD). Appendix Figure A29 shows that the specific activities with the largest point

[18]In Appendix Table A5 we report baseline means of our time use variables. The mean of self-reported Facebook minutes at baseline is 74.5 minutes per day, and the mean of reported minutes using the Facebook mobile app at baseline is 60 minutes per day.

[19]Subjects could continue using such apps and services provided they were logged in before deactivation, but they could not log in to the services as this would re-activate their accounts. They could also set up non-Facebook login credentials on other services.

estimates are going out to dinner, getting together with friends, and spending time with parents. The point estimates for the other offline activities we measure (going to the cinema, talking to friends on the phone, going to a party, going shopping, and spending time with your kids) are all very close to zero. Notwithstanding the positive effects on *offline activities*, there are no statistically significant effects on the number of friends that participants listed as having met in person last week, or on *diverse interactions* (whether or not they interacted with someone who voted differently in the last presidential election or interacted with someone from another country). We find no effects on the social interaction index, although the point estimate is positive.

The bottom group of outcomes in Figure 2 measures news consumption. As with the substitute time uses, the endline survey asked participants how much time they spent getting news from each source in the last four weeks relative to what is typical for them, and "Same" is again the average answer in the Control group. As expected, Facebook deactivation substantially reduced the extent to which people said they relied on Facebook as a news source. Consistent with the time use results, the treatment group also got substantially less news from non-Facebook social media sites (0.36 SD). The point estimates for print, radio, and TV news are all positive but statistically insignificant. Facebook deactivation has a positive but insignificant effect on Twitter use. On net, deactivation reduced the total time subjects report spending consuming news by eight minutes per day, a 15 percent reduction relative to the control group mean of 52 minutes. (We present this final result as part of our analysis of news knowledge in Section 3.)

Overall, these results suggest that Facebook is a substitute for offline activities but a complement for other online activities. This suggests the possibility that Facebook could reduce subjective well-being by reducing in-person interactions, but also impose positive political externalities by increasing news knowledge. Below, we test these possibilities more directly.

## 4.2 Effects on News and Political Outcomes

Figure 3 presents treatment effects on news and political outcomes: news knowledge, political engagement, and political polarization. News knowledge and political engagement are of interest because well-functioning democratic societies fundamentally rely on well-informed voters who actually show up to the polls to vote. Political polarization is of interest because it is may make democratic decision making less efficient, and may lead citizens to perceive democratic outcomes as less legitimate (Iyengar, Sood, and Lelkes 2012; Iyengar and Westwood 2015).

Deactivation caused substantial reductions in both self-reported attention to news and directly measured news knowledge. The top three rows show that deactivation reduced how much people reported they followed news about politics and about President Trump (by 0.14 and 0.11 SD, respectively), as well as the average minutes per day spent consuming news (a drop of eight minutes per day, or 15 percent of the control group mean). Accuracy on our news knowledge quiz falls by

0.12 standard deviations.[20] Tangibly, the Control group answered an average of 7.26 out of the 10 news knowledge questions correctly (counting "unsure" as 1/2 correct), and deactivation reduced this average by 0.14. There is no detectable effect on fake news knowledge, possibly reflecting the limited reach of even the highly shared fake news items included in our survey. Overall, deactivation reduced the news knowledge index by about 0.19 standard deviations.

There are no statistically detectable effects on political engagement. The point estimate suggests that deactivation reduced turnout by three percentage points, but the unadjusted p-value reported in Appendix Table A7 is 0.18. Similarly, the Treatment and Control groups are statistically equally likely to have clicked on any link in the post-endline politics email. Appendix Figure A34 does show a marginally significant negative effect on *voted Republican*, suggesting that deactivation may have reduced support for Republican congressional candidates. The unadjusted p-value is 0.06, the sharpened FDR-adjusted p-value is 0.08, and we had labeled this as a "secondary outcome" in our pre-analysis plan.

Deactivation reduced political polarization. Point estimates are negative for all polarization measures. The largest and most significant individual effect is on *other party understanding*: deactivation reduced the number of times that participants reportedly saw news that made them better understand the point of view of the other political party. Deactivation also decreased issue polarization, which Fiorina and Abrams (2008) single out as the "most direct" way of measuring polarization.[21] Appendix Table A6 shows that both of these effects are highly significant after adjusting for multiple hypothesis testing. The other measures with the largest point estimates are party anger and party affective polarization. Overall, deactivation reduced the political polarization index by about 0.16 standard deviations.[22]

Figure 4 illustrates how deactivation reduced issue polarization, by plotting the distribution of "issue opinions" for Democrats and Republicans in Treatment and Control at endline. Our *issue opinions* measure exactly parallels the *issue polarization* variable used in the regressions, except that we keep opinions on a left-to-right scale, with more negative indicating more agreement with

[20] Appendix F presents more analysis of the effects on news knowledge, including effects on each individual news knowledge and fake news knowledge question. All but one of the point estimates for the 10 news knowledge questions is negative. The news knowledge questions with the largest effects involve correctly responding that Elizabeth Warren's DNA test had revealed Native American ancestry and that Jeff Sessions had resigned at President Trump's request. There was also a statistically significant difference in knowledge about one fake news story: the Treatment group was less likely to correctly respond that Cesar Sayoc, the suspect in an act of domestic terrorism directed at critics of President Trump, was not a registered Democrat.

[21] Appendix Figure A30 presents results for each of the issue polarization questions. The issues for which deactivation caused the largest decrease in polarization were the direction of racial bias in policing and whether the Mueller investigation is biased.

[22] Like all of our outcome families, the polarization index includes a range of different outcomes with different interpretations. For example, the party anger and other party understanding variables are phrased to primarily measure exposure to polarizing content, which is conceptually different than affective polarization (feelings toward the other party) or issue polarization. Appendix Table A11 shows that the effect on the political polarization index is robust to excluding each of the seven individual component variables in turn, although the point estimate moves toward zero and the unadjusted p-value rises to 0.06 when omitting *other party understanding*.

the average Democratic opinion, and more positive indicating more agreement with the average Republican opinion. (By contrast, the *issue polarization* variable multiplies Democrats' responses by -1, so that a more positive value reflects more agreement with the average opinion in one's political party.[23]) We then normalize *issue opinions* to have a standard deviation of one in the Control group. The figure shows that deactivation moves both Democrats and Republicans visibly toward the center. In the Control group, the issue opinions of the average Democrat and the average Republican differ by 1.47 standard deviations. In the Treatment group, this difference is 1.35 standard deviations—about eight percent less.

Are these polarization effects large or small? As one benchmark, we can compare these effects to the increase in political polarization in the U.S. since 1996, well before the advent of social media. Using data from the American National Election Studies, Boxell (2018) calculates that the change in a different index of polarization measures increased by 0.38 standard deviations between 1996 and 2016. The 0.16 standard deviation effect of Facebook deactivation on political polarization in our sample is about 42 percent as large as this increase.[24]

Overall, these results suggest that Facebook may play a role in helping people stay informed about current events, but also increases polarization, particularly of views on political issues.

### 4.3 Effects on Subjective Well-Being

Figure 5 presents estimates of effects on subjective well-being (SWB). These outcomes are of interest because, as discussed in the introduction, many studies show cross-sectional or time-series correlations between social media use and well-being, and on this basis researchers have speculated that social media may have serious adverse effects on mental health. The outcomes are re-signed so that more positive represents better SWB—for example, the "depressed" variable is multiplied by (-1).

We find that deactivation indeed significantly increases SWB. All but one of the ten point estimates are positive. The magnitudes are relatively small overall, with the largest and most significant effects on life satisfaction (0.12 SD), anxiety (0.10 SD), depression (0.09 SD), and happiness (0.08 SD). All of these effects remain significant after adjusting for multiple hypothesis testing. The text message based measures of happiness are not significantly different from zero, with positive

[23] See footnote 12 for the definition of *issue polarization*.

[24] Specifically, Boxell's polarization index increased by 0.269 units from 1996-2016, and the standard deviation of Boxell's polarization index across people in 2016 is 0.710 units, so political polarization increased by $0.269/0.71 \approx 0.379$ standard deviations over that period. Of course, this benchmarking exercise does not imply that political polarization in the US would have increased by one-third less in the absence of Facebook, for many reasons. For example, the treatment effects in our sample from a four-week deactivation are unlikely to generalize to the US population over Facebook's 15-year life. Furthermore, some of our polarization measures are unique to our study. The one measure that appears in both Boxell's index and our index, *party affective polarization*, rose by 0.18 standard deviations between 1996 and 2016. Our point estimate of -0.06 standard deviations is about one-third of this amount, although this estimate is not statistically different from zero.

point estimates ranging from 0.01 SD to 0.06 SD. Deactivation improved our overall SWB index by 0.09 standard deviations.

Are these subjective well-being effects large or small? As one benchmark, we can consider the effect sizes in their original units, focusing on the measures with the largest effects. *Happiness* is the average response to two questions (for example, "Over the last 4 weeks, I think I was ...") on a scale from 1 (not a very happy person) to 7 (a very happy person). The Control group endline average is 4.47 out of a possible 7, and deactivation caused an average increase of 0.12. *Life satisfaction* is the extent of agreement with three questions (for example, "During the past four weeks, I was satisfied with my life") on seven-point Likert scales from "strongly disagree" "Strongly agree." The Control group endline average is 12.26 out of a possible 21, and deactivation caused an average increase of 0.56. *Depressed* and *anxious* are responses to the question, "Please tell us how much of the time during the past four weeks you felt [depressed/anxious]," where 1 is "None or almost none of the time" and 4 is "All or almost all of the time." The average responses are 2.99 and 2.60, respectively, and deactivation caused average increases of 0.08 and 0.09.

As a second benchmark, a meta-analysis of 39 randomized evaluations finds that positive psychology interventions (i.e. self-help therapy, group training, and individual therapy) improve subjective well-being (excluding depression) by 0.34 standard deviations and reduce depression by 0.23 standard deviations (Bolier et al. 2013). Thus, deactivating Facebook increased our subjective well-being index by about 25-40 percent as much as standard psychological interventions.

As a third benchmark, if we regress our baseline SWB index on key demographics (income, college completion, gender, race, age, and political party), income has one of the largest t-statistics, second only to age. As we show in Appendix Table A12, college completion is conditionally associated with 0.23 standard deviations higher SWB. Thus, the effect of deactivating Facebook is just over one-third of the conditional difference in subjective well-being between college grads and everyone else. Appendix Table A12 also shows that a $10,000 increase in income is conditionally associated with a 0.027 standard deviation increase in subjective well-being. Thus, the effect of deactivating Facebook is equal to the conditional difference in subjective well-being from about $30,000 additional income. This income equivalent is large because "money doesn't buy happiness": although income is correlated with SWB, the slope of that relationship is not very steep.

Appendix Figure A31 presents effects on the SMS outcomes by week of the experiment, to test whether the effects might have some trend over time. None of the effects on any of the three outcomes is statistically significant in any of the four weeks. The point estimates do not systematically increase or decrease over time, and if anything, the point estimates are largest in the first week. This suggests that the effects of a longer deactivation might not be different.

We can also compare our SWB effects to what we would have obtained using the kind of correlational approach taken by many previous non-experimental studies. These studies often have specific designs and outcomes that don't map closely to our paper, so it is difficult to directly compare effect sizes with other papers. We can, however, replicate the empirical strategy of simple correlation studies in our data, and compare our cross-sectional correlations to the experimental results. To do this, we regress SWB outcomes at baseline on daily average Facebook use over the past four weeks as of baseline, divided by the LATE of deactivation on daily average Facebook use between midline and endline, so that the coefficients are both in units of average use per day over the past four weeks.[25]

The baseline correlation between our SWB index and Facebook use is about three times larger than the experimental estimate of the treatment effect of deactivation (about 0.23 SD compared to 0.09 SD), and the point estimates are highly statistically significantly different. Controlling for basic demographics brings down the non-experimental estimate somewhat, but it remains economically and statistically larger than our experimental estimate. Appendix Figure A32 presents the full results for all SWB outcomes.[26] These findings are consistent with reverse causality, for example if people who are lonely or depressed spending more time on Facebook, or with omitted variables, for example if lower socio-economic status is associated with both heavy use and lower well-being. They could also reflect a difference between the relatively short-term effects measured in our experiment and the longer-term effects picked up in the cross-section. However, the lack of a detectable trend in treatment effects on the text message outcomes over the course of our experiment (as noted above and seen in Appendix Figure A31) points away from this hypothesis.

Subjects own descriptions in follow-up interviews and free-response questions are consistent with these quantitative findings, while also highlighting substantial heterogeneity in the effects. Many participants described deactivation as an unambiguously positive experience. One said in an interview,

> I was way less stressed. I wasn't attached to my phone as much as I was before. And I found I didn't really care so much about things that were happening [online] because I was more focused on my own life... I felt more content. I think I was in a better mood generally. I thought I would miss seeing everyone's day-to-day activities... I really didn't miss it at all.

A second wrote, "I realized how much time I was wasting. I now have time for other things. I've been reading books and playing the piano which I used to do daily until the phone took over." A

[25] Specifically, the non-experimental estimates are from the following regression:

$$Y_i^b = \tau \tilde{H}_i + \beta \boldsymbol{X}_i + \epsilon_i,$$

where $Y_i^b$ is participant $i$'s value of some outcome measured in the baseline survey, $\boldsymbol{X}_i$ is a vector of basic demographic variables (household income, age, and college, male, white, Republican, and Democrat indicators), and $\tilde{H}_i$ is baseline average daily Facebook use over the past four weeks (winsorized at 120 minutes per day) divided by the local average treatment effect on average daily Facebook use between midline and endline.

[26] One could also do similar experimental vs. non-experimental comparisons for other outcomes, but we have done this only for SWB because this is the main focus of the non-experimental literature in this area.

third wrote, "I realized I was using it too much and it wasn't making me happy. I hate all of the interactions I had with people in comment sections."

Many others highlighted ways in which deactivation was difficult. One said in an interview,

> I was shut off from those [online] conversations, or just from being an observer of what people are doing or thinking... I didn't like it at first at all, I felt very cut off from people that I like... I didn't like it because I spend a lot of time by myself anyway, I'm kind of an introvert, so I use Facebook in a social aspect in a very big way.

Others described the difficulty of not being able to post for special event such as family birthdays and not being able to participate in online groups groups.

Overall, our data suggest that Facebook does indeed have adverse effects on SWB. However, the magnitude of these effects is moderate and may be smaller than correlation studies would suggest, and our qualitative interviews suggest that the average effect likely masks substantial heterogeneity.

### 4.4 Post-Experiment Facebook Use and Opinions

Figure 6 presents effects of deactivation on post-experiment demand for Facebook as well as participants' subjective opinions about Facebook. These results are closely related to the findings on SWB, as we might expect participants who found deactivation increased their happiness would choose to use Facebook less in the future. They also speak more directly to the popular debate over whether social media are addictive. If deactivation reduces post-experiment Facebook use, this is consistent with standard models of addiction, such as Becker and Murphy (1988).[27]

Deactivation clearly reduced demand for Facebook. These effects are very stark, with by far the largest magnitude of any of our main findings. The effect on reported intentions to use Facebook as of the endline survey is a reduction of 0.78 standard deviations: while the average Control group participant planned to reduce future Facebook use by 22 percent, deactivation caused the Treatment group to plan to reduce Facebook use by an additional 21 percent relative to Control. In our post-endline survey a month after the experiment ended, we measured whether people actually followed through on these intentions, by asking people how much time they had spent on the Facebook mobile app on the average day in the past week. Deactivation reduces this post-endline Facebook mobile app use by 12 minutes per day, or 0.31 standard deviations. This is a 23 percent reduction relative to the Control group mean of 53 minutes per day, lining up almost exactly with the planned reductions reported at endline. However, Appendix Table A9 shows that the reduction is less than half as large (9 percent of the Control group mean) and not statistically significant (with a t-statistic of -1.40) if we limit the sample to iPhone users who reported their usage as recorded by

[27] Appendix Figure A33 presents histograms of participants' opinions about Facebook at baseline. People are heavily divided on whether Facebook is good or bad for themselves and for society and whether Facebook makes people more or less happy. Consistent with our results, people tend to think that Facebook helps people to follow the news better and makes people more politically polarized.

their Settings app, thereby excluding participants who were reporting personal estimates of their usage.

As a different (and non-self-reported) measure of post-experiment use, we can look at the speed with which users reactivated their Facebook accounts following the 24-hour post-endline period in which both Control and Treatment were deactivated. Figure 7 presents the share of our deactivation checks in which the Treatment and Control groups (with WTA less than $102) were deactivated, by day of the experiment.[28] By day 35, one week after the end of the experiment, 11 percent of the Treatment group was still deactivated, compared to three percent of the Control group. By day 91, nine weeks after the end of the experiment, five percent of the Treatment group was still deactivated, against two percent of Control. As Figure 6 shows, the local average treatment effect on the speed of reactivation is a highly significant 0.60 standard deviations. Overall, deactivation clearly decreased post-experiment use, reducing the index by 0.61 standard deviations. This is consistent with economic models of addiction, such as Becker and Murphy (1988), in which reduced consumption in one period reduces the marginal utility of consumption in future periods.

The bottom group of outcomes in Figure 6 supplement the post-experiment use outcomes by measuring participants' qualitative opinions about Facebook. These are re-signed so that more positive means more positive opinions, so agreement with the statement that "Facebook exposes people to clickbait or false news stories" and the length of text about Facebook's negative impacts are both multiplied by (-1). The results are mixed. Deactivation increases the extent to which participants think Facebook helps them follow the news better, and it also makes participants agree more that people would miss Facebook if they stopped using it. On the other hand, participants who deactivated for four weeks instead of 24 hours were more likely to say that their deactivation was good for them.[29] Deactivation increases both the *positive impacts* and *negative impacts* variables, i.e. it makes people write more about both positive and negative aspects of Facebook. Overall, deactivation had no statistically significant effect on the Facebook opinions index.

Focusing on average treatment effects risks obscuring information on the distribution of responses. Figure 8 presents the distribution of Treatment and Control responses to four key outcomes that illustrate the intensive margin of post-experiment use and key opinions about Facebook. The top left panel shows that as of the endline survey, both Treatment and Control plan to use Facebook less after the study, and deactivation makes people much more likely to plan to reduce usage by 50 percent or more (which we code at -0.75) or quit entirely (which we code at -1). The

[28]There is a slight dip in deactivation rates for the Treatment group seven days after the deactivation period began. This was caused by the fact that some participants failed to turn off a default setting in which Facebook reactivates users' profiles after seven days of deactivation. For technical reasons, our deactivation checking algorithm checked the entire Control group once every few days between midline and endline in order to check the Treatment group four times per day. After endline, we returned to checking all participants approximately once per day.

[29]One should be cautious in interpreting this effect, as it could result both from a change of opinion about Facebook and from the difference in length of the deactivation they were evaluating. As we shall see below, the Control group also tends to believe that deactivation was good for them, but the modal answer was 0 (i.e., neither good nor bad), suggesting that many people were indifferent to such a short deactivation.

top right panel presents the actual post-study usage change in percent terms: (*Facebook mobile app use* from post-endline – *Facebook mobile app use* from midline) / *Facebook mobile app use* from midline. Both Treatment and Control did indeed report using the Facebook mobile app less after the experiment than they had before, and the Treatment group reduced even more than Control.

The bottom left panel shows that both Treatment and Control tend to agree that "if people spent less time on Facebook, they would soon realize that they don't miss it," but deactivation weakened that view. On this figure, the Treatment group's average response on the scale from -5 to +5 was -1.8, while the Control group's average response is -2.0. Deactivation also appears to widen the distribution of answers, with more people strongly agreeing *and* more people strongly disagreeing. The bottom right panel shows that both Treatment and Control tend to think that deactivation was good for them, but the Treatment group is more likely to think that their four-week deactivation was good for them. On this figure, the Treatment group average response on the scale from -5 to +5 is -2.3, while the Control group's average response is -1.9. Here again, the Treatment group has a wider dispersion of responses, with more people reporting that deactivation was good for them *and* more people reporting that it was bad for them. These results highlight the importance of testing for treatment effect heterogeneity, as we will do in the next section.

To give a richer sense of how deactivation affected Facebook use, the post-endline survey included a free-response question in which we asked people to write how they had changed their Facebook use since participating in the study. We then use standard text analysis tools to determine how the Treatment and Control groups responded differently. Specifically, we processed the text by stemming words to their linguistic roots (for example, "changes," "changing," and "changed" all become "chang"), removing common "stop words" (such as "the" and "that"), and making lists of all one, two, three, and four word phrases that appeared five or more times in the sample. We then constructed Pearson's $\chi^2$ statistic, which measures the extent of differential usage rates between Treatment and Control; the phrases with the highest $\chi^2$ are especially unbalanced between the two groups. This parallels Gentzkow and Shapiro's (2010) approach to determining which phrases are used more by Republicans vs. Democrats, except we determine which phrases are used more by Treatment vs. Control.

The two panels of Table 4 present the 20 highest-$\chi^2$ phrases that were more common in Treatment and in Control. The Treatment group was relatively likely to write that they were using Facebook less or not at all ("use much less," "not use facebook anymor," "stop use facebook") or more judiciously—the phrase "use news app" is mostly from people saying that they have switched to getting news from their phone's news app instead of Facebook. By contrast, while a few of the Control group's most common phrases indicate lower use (variants of "more aware much time spend" and "use facebook slightli less"), the great majority of their relatively common phrases indicate that their Facebook use has not changed.

## 4.5 Heterogeneous Treatment Effects

In our pre-analysis plan, we specified that we would present separate estimates for subgroups defined by four primary moderators. Figure 9 presents those estimates. The top panel presents estimates for *heavy users* vs. *light users*—that is, people whose baseline reported Facebook use was above vs. below median. There is no consistent evidence that the effects are different for people who report being heavier users, perhaps because Facebook use is measured with noise.

The second panel presents estimates for *heavy news users* vs. *light news users*—that is, those who get news from Facebook fairly often or very often vs. never, hardly ever, or sometimes. As one might expect, the estimated effects for news knowledge are larger for people who get more news from Facebook, but this difference is not statistically significant. The pre-analysis plan specified that we would limit these tests to only the news and political outcomes in Section 4.2,

The third panel presents separate estimates for *active users vs. passive users.* We measure this using two questions: share of active vs. passive browsing using a question based on the Passive and Active Facebook Use Measure (Gerson, Plagnol, and Corr 2017), and "what share of your time on Facebook do you spend interacting one-on-one with people you care about." Active vs. passive users are defined as having above- vs. below-median sum of their two responses to these questions. This moderator is of interest because of a set of papers cited in the introduction suggesting that passive Facebook use can be harmful to subjective well-being, while active use might be neutral or beneficial. Perhaps surprisingly, we see no differences in the effects of deactivation on the subjective well-being index. The pre-analysis plan specified that we would limit these tests to the four families reported in the figure.

Finally, the fourth panel presents separate estimates of effects on subjective well-being text message surveys for text messages sent during the time of day when the respondent reported using Facebook the most. We see no clear differences in the effects on subjective well-being.

The pre-analysis plan also specified two secondary moderators: age (for all outcomes) and political party (limited to the news and political outcomes). We considered these secondary because we did not have a strong prior that we would be able to detect heterogeneous effects. Appendix Figure A9 presents estimates of effects on these outcomes. There are no systematic patterns.

Appendix Figure A9 also includes heterogeneity by above vs. below-median valuation of Facebook. While we added this moderator only after the pre-analysis plan was submitted, it is important because our impact evaluation sample only includes participants with WTA less than $102. Under the assumption that marginal treatment effects are monotonic in WTA, treatment effect heterogeneity within our impact evaluation sample would be informative about treatment effects for the full population. The effects for above- vs. below-median WTA differ statistically for only one index: the effects on political polarization are driven by above-median WTA participants. The above-median WTA point estimate is larger and statistically indistinguishable for two indices, smaller and statistically indistinguishable for four indices, and opposite-signed for the final index. This provides little support for the hypothesis that the effect sizes could be larger for the full Facebook user population including users with higher valuations.

Appendix E presents heterogeneous treatment effects on each individual outcome.

# 5 Measuring the Consumer Surplus from Facebook

In this section, we study participants' willingness-to-accept to deactivate Facebook, and the implications for measuring consumer surplus from zero-price digital technologies such as social media. We begin with a stylized model, and then present results.

## 5.1 Model

### Setup

We consider heterogeneous consumers who choose whether or not to consume Facebook in a stylized finite horizon dynamic problem. The utility from consuming Facebook changes as a function of prior use (*state dependence*), and current users may have incorrect beliefs about how their utility would change following deactivation (*misprediction*). State dependence will be a reduced form statistic capturing various mechanisms through which past consumption affects future demand, including habit formation, fixed costs of deactivation, or costs of deactivation that are convex in the length of deactivation. Misprediction will be a reduced form statistic capturing mechanisms causing people to misforecast how past consumption will affect future demand, including projection bias and learning.

We focus on state dependence and misprediction because these concepts capture key parts of the popular discussion around Facebook. Alter (2018), Newport (2019), many popular media articles (e.g. Ciaccia 2017; Oremus 2017), and organizations such as the Center for Humane Technology and Time to Log Off argue that Facebook and other digital technologies can be harmful and addictive. The Time to Log Off website (www.itstimetologoff.org) argues that "everyone is spending too much time on their screens" and runs "digital detox campaigns." Sagioglu and Greitemeyer (2014) document an "affective forecasting error" consistent with misprediction: people predicted that spending 20 minutes on Facebook would make them feel better, but a treatment group randomly assigned to 20 minutes of Facebook browsing actually reported feeling worse. Results from our baseline survey are consistent with this view: two-thirds of people agreed at least somewhat that "if people spent less time on Facebook, they would soon realize that they don't miss it."

We use these two "reduced form" concepts instead of "structural" behavioral models in order to map directly to our experimental design. As introduced earlier, we define the four weeks after the midline survey as month 1 and the four weeks after endline as month 2, and we index these as $t = 1$ and $t = 2$, respectively. The midline occurred just before month 1, and the endline just before month 2. State dependence would imply that the Treatment group values Facebook differently for month 2 than they did for month 1, and they know at midline that this will happen. Misprediction would imply that the Treatment group values Facebook for month 2 differently at endline than they thought they would as of midline. The Control group allows us to control for month-to-month changes in valuations, for example because Facebook might be more valuable to people during elections or holidays.

Formally, heterogeneous consumers derive utility in period $t$ from composite good consumption $c_t$ and whether or not they use Facebook, $f_t \in \{0, 1\}$. The utility function is

$$u(c_t, f_t; f_{t-1}) = c_t + (\phi f_{t-1} + \xi + \omega_t) f_t, \tag{3}$$

where $\xi$ is the utility from Facebook in the average period, $\omega_t$ is a period-specific utility shock that is revealed at the beginning of period $t$ and has expectation zero across periods ($\mathbb{E}_t[\omega_t] = 0$), and $\phi$ represents state dependence. $\phi > 0$ could reflect habit formation or a fixed cost of deactivation, while $\phi < 0$ could reflect convex costs of deactivation.

At the beginning of period $\tau$, consumers predict their utility for that period and for future periods. Predicted utility for $t \geq \tau$ is

$$\tilde{u}(c_t, f_t; f_{t-1}, f_{\tau-1}) = c_t + (\phi f_{t-1} + \alpha f_{\tau-1} + \xi + \omega_t) f_t. \tag{4}$$

If consumers used Facebook in the most recent period, they overestimate utility from Facebook by amount $\alpha$. This timing maps to the timing of our experiment, where valuations for month $t$ are elicited at the end of month $t - 1$ and all subjects are current users of Facebook at that point in time. $\alpha \neq 0$ could result from projection bias or learning—tangibly, if re-experiencing life without Facebook causes people to unexpectedly realize that they value Facebook more or less than they had thought.

At the beginning of period $\tau$, consumers maximize predicted utility subject to a budget constraint $W$:

$$\tilde{U}_\tau(\cdot; f_{\tau-1}) = \sum_{t=\tau}^{T} \tilde{u}(c_t, f_t; f_{t-1}, f_{\tau-1}), \quad s.t. \sum_{t=\tau}^{T} c_t \leq W. \tag{5}$$

We assume that in the absence of our treatment people predict that they will always use Facebook in future periods after the current period: $f_t = 1, \forall t > \tau$. This is approximately true in our data, and it substantially simplifies the dynamic problem.

### Valuation of Facebook

We can now write the three WTAs elicited in our experiment in terms of the model. We denote $v_{t,\tau}(f_{t-1}, f_{\tau-1})$ as the valuation of Facebook for period $t$ as evaluated at the beginning of period

$\tau$, conditional on $f_{t-1}$ and $f_{\tau-1}$. Consumers determine $v_{t,\tau}(f_{t-1}, f_{\tau-1})$ by the payment that would be required to equalize $\tilde{U}_\tau(f_t = 0; f_{\tau-1})$ with $\tilde{U}_\tau(f_t = 1; f_{\tau-1})$. See Appendix H.1 for derivations. The WTA for month $t = 1$ deactivation just before the beginning of month 1 is

$$v_{1,1}(1,1) = \underbrace{\phi + \alpha + \xi + \omega_1}_{\text{t=1 Facebook utility loss}} + \underbrace{\phi}_{\text{t=2 Facebook utility loss}}. \tag{6}$$

The tabular display below presents the WTAs for month 2 deactivation for Treatment and Control as of midline and endline. We define $\Delta v := v_{2,2}(f_1, f_1) - v_{2,1}(f_1, 1)$ as the difference between WTA for month 2 deactivation as of endline vs. as of midline. For the Control group, $\Delta v$ results from the time effect $\omega_2$. For the Treatment group, $\Delta v$ results from the time effect and misprediction: $\omega_2 - \alpha$. The difference in differences is thus $-\alpha$, our measure of misprediction. The display also shows that $\phi$ can be estimated from the Treatment – Control difference in month 2 valuation as of midline. The Treatment – Control difference in month 2 valuation as of endline is $-\alpha - \phi$.

**Willingness-to-Accept to Deactivate Facebook in Month 2**

| | | *Group* | | |
|---|---|---|---|---|
| | | Treatment | Control | Difference |
| *Time of elicitation* | Midline | $\alpha + \xi + \phi$ | $\phi + \alpha + \xi + \phi$ | $-\phi$ |
| | Endline | $\xi + \omega_2 + \phi$ | $\phi + \alpha + \xi + \omega_2 + \phi$ | $-\alpha - \phi$ |
| | Difference | $\omega_2 - \alpha$ | $\omega_2$ | $-\alpha$ |

### Consumer Surplus

Misprediction creates ambiguity in the calculation of "true" consumer surplus. Should we use people's predicted valuations or their actual valuations after experiencing deactivation? In the language of Bernheim and Rangel (2009), it is unclear which valuations to consider as the "welfare-relevant domain." Models of projection bias such as Loewenstein, O'Donoghue, and Rabin (2003) typically frame the updated valuation as "true" utility. Our model follows this approach by allowing "predicted" utility $\tilde{u}$ to differ from true utility by amount $\alpha f_{\tau-1}$.[30]

In the model, the consumer surplus from Facebook over $T$ periods is

$$CS = \sum_{t=1}^{T} \mathbb{E}u(c_t, 1; 1, 0) - \sum_{t=1}^{T} \mathbb{E}u(c_t, 0; 0, 0) = T \cdot (\xi + \phi). \tag{7}$$

[30] Loewenstein, O'Donoghue, and Rabin (2003, page 1211) write, "for instance, if current consumption has deleterious effects on future well-being, and projection bias leads the person to underappreciate these effects, she may overconsume relative to what would maximize her true intertemporal utility." We report average valuations from all three WTA elicitations for readers who differ on which is welfare-relevant or, in the spirit of Bernheim and Rangel (2009), wish to construct bounds on welfare given the ambiguity between alternative estimates.

Brynjolfsson, Eggers, and Gannamaneni (2018), Corrigan et al. (2018), Mosquera et al. (2018), and Sunstein (2019) measure willingness-to-accept to not use Facebook. In our model, willingness-to-accept to not use Facebook for $T$ periods beginning in $t = 1$, denoted $v_{\{1,\ldots,T\},1}(1,1)$, is

$$v_{\{1,\ldots,T\},1}(1,1) = \omega_1 + T(\phi + \alpha + \xi). \tag{8}$$

.

In our model, this WTA $v_{\{1,\ldots,T\},1}(1,1)$ differs from actual consumer surplus over $T$ periods by the following ratio:

$$\frac{v_{\{1,\ldots,T\},1}(1,1) - CS}{v_{\{1,\ldots,T\},1}(1,1)} = \frac{\omega_1 + T\alpha}{v_{\{1,\ldots,T\},1}(1,1)}. \tag{9}$$

Thus, in our model, WTA to deactivate differs from consumer surplus for two reasons: the time effect $\omega_1$ and misprediction $\alpha$.

### 5.2 Results

In this section, we first present consumer surplus estimates based on month 1 WTA. We then adjust those estimates through the lens of our model to account for misprediction.

People in our sample derive substantial value from Facebook. Figure 10 presents the histogram of willingness-to-accept to deactivate Facebook for the four weeks after midline instead of only the 24 hours after midline. The median is \$100, and almost 20 percent had valuations greater than \$500. After winsorizing valuations at \$1000, the mean is \$203. After re-weighting the sample to match the observable characteristics of Facebook users in Table 2, the median is still \$100, and the winsorized mean is \$180. As a benchmark, a flow of \$100 (\$180) in consumer surplus per 27 days would imply an annual flow of \$1,351 (\$2,433) per user. Multiplying this by the estimated 172 million US Facebook users would give a nationwide annual consumer surplus flow of \$230 billion (\$420 billion).

Our estimated valuations per day of Facebook abstention are larger than most, but not all, other estimates. In an online panel weighted for national representativeness, Brynjolfsson, Eggers, and Gannamaneni (2018) estimate that the mean WTA to not use Facebook for one month is \$48, and that the median WTA to hypothetically stop using social media for one year was \$205 in 2016 and \$322 in 2017. In their sample of European college students, Brynjolfsson, Eggers, and Gannamaneni (2018) find a median WTA of \$175 for one month.[31] In samples of college students, residents of a college town, and Amazon MTurk workers, Corrigan et al. (2018) estimate that the mean annualized WTA to deactivate Facebook ranges from \$1,139 to \$1,921, depending on the sample and the length of deactivation. In a sample of college students, Mosquera et al.

[31] Appendix Figure A35 compares our demand curve to the Brynjolfsson, Eggers, and Gannamaneni (2018) demand curves.

(2018) estimate that the median (mean) WTA to not use Facebook for one week is \$15 (\$25). In an unincentivized (stated preference) survey of MTurk workers, Sunstein (2019) found a median willingness-to-pay for Facebook of \$1 per month, and a median willingness-to-accept to not use Facebook of \$59 per month.

There are several reasons why our valuations might differ from these other studies. First, we and Corrigan et al. told participants that they would need to "deactivate" their accounts and described how that deactivation would be enforced. By contrast, Brynjolfsson et al. and Mosquera et al. told participants that they would need to "not use" their accounts, and Mosquera et al. did not describe how this would be enforced.[32] WTA could naturally be lower for an unenforced deactivation than for an enforced deactivation. Furthermore, staying deactivated (as opposed to simply not logging in) also requires users to not log into other apps such as Spotify using their Facebook login credentials, and our instructions emphasized this. Second, each study required Facebook abstention over different lengths of time, and multiple studies present evidence that the length of abstention affects the average per-day valuation. Third, the samples differ. We screened out people who reported using Facebook 15 minutes or less per day, and while we re-weight the average WTAs to match the average observables of Facebook users (including average daily usage), this re-weighting may implicitly overstate the WTA of people who don't use Facebook very much.

Figure 11 presents the average of WTA in Treatment and Control in each of the three elicitations: at midline for month 1 deactivation, at midline for month 2 deactivation, and at endline for month 2 deactivation. Recall that Treatment and Control include only those with $v_{1,1}(1,1)_i < \$102$. Because of outliers in the month 2 WTAs, we must winsorize WTA. We winsorize at \$170 for our primary estimates, as this is the upper bound of the distribution of BDM offers that we actually made for deactivation. The Treatment group's month 2 valuation jumps substantially relative to its post-midline valuation, while the Control group's month 2 valuation does not. In the context of our model, this would imply large state dependence $\phi$, although we believe that this could be biased upwards for reasons discussed below.

Figure 11 also illustrates $\Delta v$: the change in month 2 valuation between midline and endline. The Control group's $\Delta v$ is slightly positive, which in our model captures a time effect $\omega_2$. In open-answer questions, some people wrote that they were less willing to deactivate during the Thanksgiving holiday, and they may not have foreseen this as of the midline survey on October 11th. By contrast, the Treatment group's $\Delta v$ is slightly negative. Thus, the treatment effect on $\Delta v$ is somewhat negative. In our model, this is interpreted as the Treatment group slightly over-predicting their valuation of Facebook.

[32] Brynjolfsson et al.'s WTA elicitation stated that the experimenters "will randomly pick 1 out of every 200 respondents and her/his selection will be fulfilled," and that they could enforce deactivation by observing subjects' time of last login, "given your permission." In practice, the deactivation was mostly not enforced: of the ten subjects randomly selected for enforcement, one gave permission. Mosquera et al. told participants that they would "require" that they "not use their Facebook accounts."

We can estimate $\phi$ and $\alpha$ using the following regression:

$$Y_i = \tau D_i + \rho v_{1,1}(1,1)_i + \nu_s + \varepsilon_i, \tag{10}$$

instrumenting for $D_i$ with $T_i$. This parallels Equation (1). When we set $Y_i = v_{2,1}(f_1,1)_i$, $\hat{\tau}$ is our estimate of $-\phi$. When we set $Y_i = \Delta v_i$, $\hat{\tau}$ is our estimate of $-\alpha$.

Table 5 presents results, winsorizing all WTAs at \$170 in columns 1 and 3, and at \$1,000 in columns 2 and 4. As we saw in Figure 10, month 2 WTA as of midline is about \$60 to \$90 larger in Treatment compared to Control, suggesting a large positive $\phi$. Relative to the Control group, the Treatment group reduced its post-endline valuation by \$14 to \$18. This suggests slight over-prediction equal to 13 percent of the Treatment group's average month 2 WTA elicited at midline, which is \$103 (\$136) when winsorizing at \$170 (\$1,000). Applying this to the above consumer surplus estimate suggests a misprediction-adjusted estimate of $\$420 \times (1 - 0.13) \approx \$365$ billion per year.

### 5.3 Caveats

We used open-answer questions in the post-endline survey and qualitative interviews to understand why the WTA participants reported for month 2 was so much larger than the WTA they reported for month 1. This additional information suggests that some of the large gap is due to costs of deactivation being convex in the length of deactivation, consistent with our model's interpretation of this fact as evidence for a large negative state dependence parameter $\phi$. In post-endline open answer survey questions, some people in the Treatment group wrote that they were much less comfortable deactivating for eight weeks instead of four, as they would have to make much more extensive arrangements to communicate with friends, co-workers, and schoolmates during a longer deactivation. However, our additional survey work suggests that some of the Treatment group's increased valuation was driven not by valuations of Facebook *per se*, but instead by factors such as anchoring on the \$102 BDM payment that they had been offered between the month 1 elicitation and the month 2 elicitation during the midline survey. Such anchoring is consistent with the results of Bohm, Lindén, and Sonnegård (1997) and Mazar, Koszegi, and Ariely (2014). On net, these results make us cautious in interpreting $\phi$, and we think more research is required to obtain a precise estimate of state dependence.

If anchoring has the same effects on the Treatment group's month 2 WTA at endline that it had at midline, then our estimate of misprediction $\alpha$ is unbiased. If the anchoring effects decay somewhat, this would bias $\alpha$ away from zero, which would imply that the true $\alpha$ is less than our estimate of 13 percent of WTA. This would further strengthen the headline result that misprediction represents only a small share of valuations, and would reinforce the conclusion that Facebook generates substantial consumer surplus, perhaps in the hundreds of billions of dollars each year in the United States alone.

There are a number of additional caveats to this calculation. First, while we re-weight the sample to match the average Facebook user on daily usage time and other observables, our sample is not representative on some observables and may not be representative on relevant unobservables. Second, we estimate valuation of Facebook holding networks fixed. Due to network externalities, valuations could be quite different if participants' friends and family also deactivated. Third, we consider WTA for only 27-day periods, and our quantitative and qualitative evidence suggests at least some non-linearity in the costs of deactivation per unit time. Fourth, the dynamics in our model are over-simplified due to our two-period experiment. In reality, habit formation likely depends on more than just last period's activity. Fifth, our model is "reduced form," in the sense that there are many possible specific behavioral models that could generate the patterns of WTA we see in the data. Sixth, we do not consider other possible behavioral biases that could affect social media consumption, such as self-control problems. If social media is a temptation good, then people will "overconsume" it relative to their own long-run preferences. Note that because we elicit WTA for *future* deactivation, we measure consumer surplus from the point of view of the "long-run self."

# 6 Conclusion

Our results leave little doubt that Facebook produces large benefits for its users. A majority of people in our sample value four weeks of access at $100 or more, and these valuations could imply annual consumer surplus gains in the hundreds of billions of dollars in the US alone. The 60 minutes our participants spend on Facebook each day is itself suggestive of the substantial value it provides. Our results on news consumption and knowledge suggest that Facebook plays an important role as a source of (real) news and information. Our participants' answers in free response questions and follow-up interviews make clear the diverse ways in which Facebook can improve people's lives, whether as a source of entertainment, a means to organize a charity or an activist group, or a vital social lifeline for those who are otherwise isolated. Any discussion of social media's downsides should not obscure the basic fact that it fulfills deep and widespread needs.

Notwithstanding, our results also make clear that the downsides are real. We offer the largest-scale experimental evidence measuring a wide set of potential impacts at both the individual and societal level. We find that four weeks without Facebook improves subjective well-being and substantially reduces post-experiment demand, suggesting that forces such as addiction and projection bias may cause people to use Facebook more than they otherwise would. We find that while deactivation makes people less informed, it also makes them less polarized by at least some measures, consistent with the concern that social media have played some role in the recent rise of polarization in the US. The estimated magnitudes imply that these negative effects are large enough to be real concerns, but also smaller in many cases than what one might have expected given prior research and popular discussion.

The trajectory of views on social media—with early optimism about great benefits giving way to alarm about possible harms—is a familiar one. Innovations from novels to TV to nuclear energy have had similar trajectories. Along with the excellent existing work by other researchers, we hope that our analysis can help move the discussion from simplistic caricatures to hard evidence, and to provide a sober assessment of the way a new technology affects both individual people and larger social institutions.

## References


Abeele, Mariek M. P. Vanden, Marjolijn L. Antheunis, Monique M. H. Pollmann, Alexander P. Schouten, Christine C. Liebrecht, Per J. van der Wijst, Marije A. A. van Amelsvoort, Jos Bartels, Emiel J. Krahmer, and Fons A. Maes. 2018. "Does Facebook Use Predict College Students' Social Capital? A Replication of Ellison, Steinfield, and Lampe's (2007) Study Using the Original and More Recent Measures of Facebook Use and Social Capital." *Communication Studies* 69 (3):272–282.

Acland, Dan and Matthew Levy. 2015. "Naivete, Projection Bias, and Habit Formation in Gym Attendance." *Management Science* 61 (1):146–160.

Allcott, Hunt and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2):211–236.

Alter, Adam. 2018. *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked*. Penguin Random House.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484):1481–1495.

Appel, Helmut, Alexander L. Gerlach, and Jan Crusius. 2016. "The Interplay Between Facebook Use, Social Comparison, Envy, and Depression." *Current Opinion in Psychology* 9:44–49.

Argyle, Michael. 2001. *The Psychology of Happiness*. Routledge.

Baker, David A. and Guillermo Perez Algorta. 2016. "The Relationship Between Online Social Networking and Depression: A Systematic Review of Quantitative Studies." *Cyberpsychology, Behavior, and Social Networking* 19 (11):638–648.

Bartels, Larry M. 1993. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87 (2):267–285.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *The American Economic Review* 81 (2):237–241.

Becker, Gary S. and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4):675–700.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak. 1964. "Measuring Utility by a Single-Response Sequential Method." *Economic Information, Decision, and Prediction* 9 (3):226–232.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika* 93 (3):491–507.

Bernheim, B. Douglas and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-theoretic Foundations for Behavioral Welfare Economics." *The Quarterly Journal of Economics* 124 (1):51–104.

Besley, Timothy and Robin Burgess. 2001. "Political Agency, Government Responsiveness and the Role of the Media." *European Economic Review* 45 (4-6):629–640.

Bohm, Peter, Johan Lindén, and Joakim Sonnegård. 1997. "Eliciting Reservation Prices: Becker–DeGroot–Marschak Mechanisms vs. Markets." *The Economic Journal* 107 (443):1079–1089.

Bolier, Linda, Merel Haverman, Gerben J. Westerhof, Heleen Riper, Filip Smit, and Ernst Bohlmeijer. 2013. "Positive Psychology Interventions: A Meta-Analysis of Randomized Controlled Studies." *BMC Public Health* 13 (1):119.

Boxell, Levi. 2018. "Demographic Change and Political Polarization in the United States." SSRN Working Paper.

Brynjolfsson, Erik, Felix Eggers, and Avinash Gannamaneni. 2018. "Using Massive Online Choice Experiments to Measure Changes in Well-being." NBER Working Paper.

Brynjolfsson, Erik and Joo Hee Oh. 2012. "The Attention Economy: Measuring the Value of Free Digital Services on the Internet." In *Proceedings of the 33rd International Conference on Information Systems*.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2018. "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics." In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Brynjolfsson, Erik and Adam Saunders. 2009. "What the GDP Gets Wrong (Why Managers Should Care)." *MIT Sloan Management Review* 51 (1):95.

Burke, Moira and Robert E. Kraut. 2014. "Growing Closer on Facebook: Changes in Tie Strength Through Social Network Site Use." In *Proceedings of the SIGCHI conference on human factors in computing systems*. 4187–4196.

———. 2016. "The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength." *Journal of Computer-Mediated Communication* 21 (4):265–281.

Burke, Moira, Robert E. Kraut, and Cameron Marlow. 2011. "Social Capital on Facebook: Differentiating Uses and Users." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 571–580.

Burke, Moira, Cameron Marlow, and Thomas Lento. 2010. "Social Network Activity and Social Well-Being." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1909–1912.

Busse, Meghan R., Devin G. Pope, Jaren C. Pope, and Jorge Silva-Risso. 2015. "The Psychological Effect of Weather on Car Purchases." *The Quarterly Journal of Economics* 130 (1):371–414.

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. 2016. "Does the United States have a productivity slowdown or a measurement problem?" *Brookings Papers on Economic Activity* 2016 (1):109–182.

Charness, Gary and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica* 77 (3):909–931.

Chopik, William J. 2017. "Associations among relational values, support, health, and well-being across the adult lifespan." *Personal Relationships* 24 (2):408–422.

Ciaccia, Chris. 2017. "Facebook, cocaine, opioids: How addictive is the social network?" Fox News. https://www.foxnews.com/tech/facebook-cocaine-opioids-how-addictive-is-the-social-network.

Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang. 2007. "Projection Bias in Catalog Orders." *American Economic Review* 97 (4):1217–1249.

Corrigan, Jay R., Saleem Alhabash, Matthew Rousu, and Sean B. Cash. 2018. "How much is social media worth? Estimating the value of Facebook by paying users to stop using it." *PloS one* 13 (12):e0207101.

Csikszentmihalyi, Mihaly and Reed Larson. 2014. "Validity and Reliability of the Experience-Sampling Method." In *Flow and the Foundations of Positive Psychology*. Springer, 35–54.

DellaVigna, Stefano and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics* 2 (1):643–669.

DellaVigna, Stefano and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics* 122 (3):1187–1234.

DellaVigna, Stefano and Eliana La Ferrara. 2015. "Economic and Social Impacts of the Media." In *Handbook of Media Economics*, vol. 1. Elsevier, 723–768.

Deters, Fenne große and Matthias R. Mehl. 2012. "Does Posting Facebook Status Updates Increase or Decrease Loneliness? An Online Social Networking Experiment." *Social Psychological and Personality Science* 4 (5):579–586.

Diener, Ed, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. "The Satisfaction With Life Scale." *Journal of Personality Assessment* 49 (1):71–75.

Ellison, Nicole B., Charles Steinfield, and Cliff Lampe. 2007. “The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites.” *Journal of Computer-Mediated Communication* 12 (4):1143–1168.

Enikolopov, Ruben and Maria Petrova. 2015. “Media Capture: Empirical Evidence.” In *Handbook of Media Economics*, vol. 1. Elsevier, 687–700.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. “Media and Political Persuasion: Evidence from Russia.” *American Economic Review* 101 (7):3253–85.

Facebook. 2016. “Facebook Q1 2016 Results - Earnings Call Transcript.” https://seekingalpha.com/article/3968783-facebook-fb-mark-elliot-zuckerberg-q1-2016-results-earnings-call-transcript.

———. 2018. “Facebook Reports Third Quarter 2018 Results.” Press Release. https://investor.fb.com/investor-news/press-release-details/2018/Facebook-Reports-Third-Quarter-2018-Results/default.aspx.

Fardouly, Jasmine and Lenny R. Vartanian. 2015. “Negative Comparisons About One’s Appearance Mediate the Relationship Between Facebook Usage and Body Image Concerns.” *Body Image* 12:82–88.

Fiorina, Morris P. and Samuel J. Abrams. 2008. “Political Polarization in the American Public.” *Annual Review Political Science* 11:563–588.

Frison, Eline and Steven Eggermont. 2015. “Toward an Integrated and Differential Approach to the Relationships Between Loneliness, Different Types of Facebook Use, and Adolescents’ Depressed Mood.” *Communication Research*. Advance online publication. doi:10.1177/0093650215617506.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. “Habit Formation in Voting: Evidence from Rainy Elections.” *American Economic Journal: Applied Economics* 8 (4):160–188.

Gallup. 2018a. “Americans’ Views of Donald Trump’s Personal Characteristics (Trends).” https://news.gallup.com/poll/235916/americans-views-donald-trump-personal-characteristics-trends.aspx.

———. 2018b. “Opposition to Kavanaugh Had Been Rising Before Accusation.” https://news.gallup.com/poll/242300/opposition-kavanaugh-rising-accusation.aspx.

———. 2018c. “Ten Takeaways About Americans’ Views of Guns.” https://news.gallup.com/opinion/polling-matters/233627/ten-takeaways-americans-views-guns.aspx.

Gentzkow, Matthew. 2006. "Television and Voter Turnout." *The Quarterly Journal of Economics* 121 (3):931–972.

———. 2016. "Polarization in 2016." Toulouse Network for Information Technology Whitepaper.

Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw. 2011. "How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *American Political Science Review* 105 (1):135–150.

Gerber, Alan S. and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (3):653–663.

Gerber, Alan S., Dean Karlan, and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1 (2):35–52.

Gerson, Jennifer, Anke C. Plagnol, and J. Corr, Philip. 2017. "Passive and Active Facebook Use Measure (PAUM): Validation and Relationship to the Reinforcement Sensitivity Theory." *Personality and Individual Differences* 117:81–90.

Gonzales, Amy L. and Jeffrey T. Hancock. 2011. "Mirror, Mirror on my Facebook Wall: Effects of Exposure to Facebook on Self-Esteem." *Cyberpsychology, Behavior, and Social Networking* 14 (1-2):79–83.

Gruber, Jonathan and Botond Köszegi. 2001. "Is Addiction "Rational"? Theory and Evidence." *The Quarterly Journal of Economics* 116 (4):1261–1303.

Howard, Philip N., Aiden Duffy, Deen Freelon, Muzammil M. Hussain, Will Mari, and Marwa Maziad. 2011. "Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?" Available at SSRN 2595096.

Huber, Gregory A. and Kevin Arceneaux. 2007. "Identifying the Persuasive Effects of Presidential Advertising." *American Journal of Political Science* 51 (4):957–977.

Hughes, Mary Elizabeth, Linda J. Waite, Louise C. Hawkley, and John T. Cacioppo. 2004. "A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies." *Res Aging* 26 (6):655–672.

Hunt, Melissa G., Rachel Marx, Courtney Lipson, and Jordyn Young. 2018. "No More FOMO: Limiting Social Media Decreases Loneliness and Depression." *Journal of Social and Clinical Psychology* 37 (10):751–768.

Huppert, Felicia A., Nic Marks, Andrew Clark, Johannes Siegrist, Alois Stutzer, Joar Vittersø, and Morten Wahrendorf. 2009. "Measuring Well-being Across Europe: Description of the ESS Well-being Module and Preliminary Findings." *Social Indicators Research* 91:301–315.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2016. "Handwashing and Habit Formation." Yale University, Economic Growth Center.

Irvine, Mark. 2018. "Facebook Ad Benchmarks for YOUR Industry." https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology: Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3):405–431.

Iyengar, Shanto and Sean J. Westwood. 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3):690–707.

Kahneman, Daniel, Alan B. Krueger, David Schkade, Norbert Schwarz, and Arthur A. Stone. 2006. "Would You Be Happier If You Were Richer? A Focusing Illusion." *Science* 312:1908–1910.

Kirkpatrick, David. 2011. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. Simon & Schuster.

Krasnova, Hannah, H. Wenninger, T. Widjaja, and Peter Buxmann. 2013. "Envy on Facebook: A Hidden Threat to Users' Life Satisfaction." In *11th International Conference on Wirtschaftsinformatik*.

Kross, Ethan, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablack, John Jonides, and Oscar Ybarra. 2013. "Facebook Use Predicts Declines in Subjective Well-being in Young Adults." *PloS one* 8 (8):e69841.

Lee, David Seungjae. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76 (3):1071–1102.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *The Quarterly Journal of Economics* 118 (4):1209–1248.

Lyubomirsky, Sonja and Heidi S. Lepper. 1999. "A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation." *Social Indicators Research* 46:137–155.

Mabe, Annalise G., K. Jean Forney, and Pamela K. Keel. 2014. "Do You "Like" My Photo? Facebook Use Maintains Eating Disorder Risk." *International Journal of Eating Disorders* 47 (5):516–523.

Marotta, Veronica and Alessandro Acquisti. 2017. "Online Distractions, Website Blockers, and Economic Productivity: A Randomized Field Experiment." Preliminary Draft.

Martin, Gregory J. and Ali Yurukoglu. 2017. "Bias in Cable News: Persuasion and Polarization." *American Economic Review* 107 (9):2565–99.

Mazar, Nina, Botond Koszegi, and Dan Ariely. 2014. "True Context-dependent Preferences? The Causes of Market-dependent Valuations." *Journal of Behavioral Decision Making* 27 (3):200–208.

Molla, Rani and Kurt Wagner. 2018. "People spend almost as much time on Instagram as they do on Facebook." Recode. https://www.recode.net/2018/6/25/17501224/instagram-facebook-snapchat-time-spent-growth-data.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie. 2018. "The Economic Effects of Facebook." Working Paper, Texas A&M University.

Myers, David G. 2000. "The Funds, Friends, and Faith of Happy People." *American Psychologist* 55 (1):56.

Nakamura, Leonard I., Jon Samuels, and Rachel H. Soloveichik. 2016. "Valuing "Free" Media in GDP: An Experimental Approach." FRB of Philadelphia Working Paper No. 16-24.

Napoli, Philip M. 2014. "Measuring Media Impact." The Norman Lear Center. http://www.learcenter.org/pdf/measuringmedia.pdf.

Newport, Cal. 2019. *Digital Minimalism: Choosing a Focused Life in a Noisy World.* Penguin Random House.

Olken, Benjamin A. 2009. "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages." *American Economic Journal: Applied Economics* 1 (4):1–33.

Oremus, Will. 2017. "Addiction for Fun and Profit." Slate. https://slate.com/technology/2017/11/facebook-was-designed-to-be-addictive-does-that-make-it-evil.html.

Pew Research Center. 2016. "On Views of Race and Inequality, Blacks and Whites Are Worlds Apart." http://www.pewsocialtrends.org/2016/06/27/on-views-of-race-and-inequality-blacks-and-whites-are-worlds-apart/.

———. 2017. "American Trends Panel, Wave 25 March, Final Topline, March 13 - March 27, 2017." http://www.journalism.org/wp-content/uploads/sites/8/2017/05/PJ_2017.05.10_media-attitudes_topline.pdf.

———. 2018a. “News Use Across Social Media Platforms 2018.” http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/.

———. 2018b. “Political Survey, Final Topline, April 25 - May 1, 2018.” http://assets.pewresearch.org/wp-content/uploads/sites/5/2018/05/10094104/05-10-18-foreign-policy-topline-for-release.pdf.

———. 2018c. “Public Is Skeptical of the Iran Agreement - and Trump’s Handling of the Issue.” http://www.people-press.org/2018/05/08/public-is-skeptical-of-the-iran-agreement-and-trumps-handling-of-the-issue/.

———. 2018d. “Sexual Harassment at Work in the Era of #MeToo.” http://www.pewsocialtrends.org/2018/04/04/sexual-harassment-at-work-in-the-era-of-metoo/.

———. 2018e. “Shifting Public Views on Legal Immigration Into the U.S.” http://www.people-press.org/2018/06/28/shifting-public-views-on-legal-immigration-into-the-u-s/.

———. 2018f. “Social Media Use in 2018.” http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/.

Reis, Harry T., W. Andrew Collins, and Ellen Berscheid. 2000. “The Relationship Context of Human Behavior and Development.” *Psychological Bulletin* 126 (6):844.

Sagioglu, Christina and Tobias Greitemeyer. 2014. “Facebook’s emotional consequences: Why Facebook causes a decrease in mood and why people still use it.” *Computers in Human Behavior* 35:359–363.

Satici, Seydi Ahmet and Recep Uysal. 2015. “Well-being and problematic Facebook use.” *Computers in Human Behavior* 49:185–190.

Settle, Jamie E. 2018. *Frenemies: How Social Media Polarizes America.* Cambridge University Press.

Shakya, Holly B. and Nicholas A. Christakis. 2017. “Association of Facebook Use With Compromised Well-being: A Longitudinal Study.” *American Journal of Epidemiology* 185 (3):203–211.

Simonsohn, Uri. 2010. “Weather to go to college.” *The Economic Journal* 120 (543):270–280.

Spenkuch, Jörg L and David Toniatti. 2016. “Political advertising and election outcomes.” CESifo Working Paper Series 5780.

Stone, Arthur A. and Saul Shiffman. 1994. “Ecological Momentary Assessment (EMA) in Behavorial Medicine.” *Annals of Behavioral Medicine* 16 (3):199–202.

Strömberg, David. 2015. "Media and Politics." *Annual Review of Economics* 7 (1):173–205.

Sunstein, Cass. 2001. *Republic.com*. Princeton University Press.

———. 2017. *Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

———. 2019. "Valuing Facebook." *Behavioural Public Policy* Forthcoming.

Syverson, Chad. 2017. "Challenges to Mismeasurement Explanations for the US Productivity Slowdown." *Journal of Economic Perspectives* 31 (2):165–86.

Tandoc, Edson C., Patrick Ferrucci, and Margaret Duffy. 2015. "Facebook use, envy, and depression among college students: Is facebooking depressing?" *Computers in Human Behavior* 43:139–146.

Theocharis, Yannis and Will Lowe. 2016. "Does Facebook increase political participation? Evidence from a field experiment." *Information, Communication, and Society* 19 (10):1465–1486.

Tromholt, Morten. 2016. "The Facebook Experiment: Quitting Facebook Leads to Higher Levels of Well-Being." *Cyberpsychology, Behavior, and Social Networking* 19 (11):661–666.

Twenge, Jean M. 2017. *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy–and Completely Unprepared for Adulthood–and What That Means for the Rest of Us*. Simon and Schuster.

Twenge, Jean M., Thomas E. Joiner, Megan L. Rogers, and Gabrielle N. Martin. 2018. "Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time." *Clinical Psychological Science* 6 (1):3–17.

Twenge, Jean M., Gabrielle N. Martin, and W. Keith Campbell. 2018. "Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology." *Emotion* 18 (6):765–780.

Twenge, Jean M. and Heejung Park. 2017. "The Decline in Adult Activities Among U.S. Adolescents, 1976–2016." *Child Development*. Advance online publication. doi:10.1111/cdev.12930.

Twenge, Jean M., Ryne A. Sherman, and Sonja Lyubomirsky. 2016. "More Happiness for Young People and Less for Mature Adults: Time Period Differences in Subjective Well-Being in the United States, 1972–2014." *Social Psychological and Personality Science* 7 (2):131–141.

Vanman, Eric J., Rosemary Baker, and Stephanie J. Tobin. 2018. "The Burden of Online Friends: The Effects of Giving Up Facebook on Stress and Well-Being." *The Journal of Social Psychology* 158 (4):496–507.

Verduyn, Phuluppe, David Seungjae Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross. 2015. "Passive Facebook Usage Undermines Affective Well-Being: Experimental and Longitudinal Evidence." *Journal of Experimental Psychology* 144 (2):480–488.

Table 1: **Sample Sizes**

| Phase | Sample size |
|---|---|
| Recruitment and baseline | N=1,690,076 shown ads |
| | N=30,064 clicked on ads |
| | N=14,324 passed pre-screen |
| | N=5,974 consented |
| | N=3,234 finished baseline |
| | N=3,051 had valid baseline, were randomized |
| Midline | N=3,015 began midline |
| | N=2,844 finished midline, of which: |
| | N=1,661 in impact evaluation sample |
| Endline | N=2,823 began endline |
| | N=2,795 finished endline, of which: |
| | N=1,639 in impact evaluation sample |
| Post-endline | N=2,137 reported Facebook mobile app use, of which: |
| | N=1,219 in impact evaluation sample |

Table 2: **Sample Demographics**

| | (1) Impact evaluation sample | (2) Facebook users | (3) US population |
|---|---|---|---|
| Income under $50,000 | 0.40 | 0.41 | 0.42 |
| College | 0.51 | 0.33 | 0.29 |
| Male | 0.43 | 0.44 | 0.49 |
| White | 0.68 | 0.73 | 0.74 |
| Age under 30 | 0.52 | 0.26 | 0.21 |
| Republican | 0.13 | | 0.26 |
| Democrat | 0.42 | | 0.20 |
| Facebook minutes | 74.52 | 45.00 | |

Notes: Column 1 presents average demographics for the impact evaluation sample: participants who were willing to accept less than $102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. Column 2 presents our estimate of average demographics of American adults with a Facebook account. The top five numbers in Column 2 are inferred from a Pew Research Center (2018f) survey of social media use by demographic group. The bottom number in Column 2 (the average of 45 minutes of Facebook use per day) is approximated on those basis of sources such as Facebook (2016) and Molla and Wagner (2018). Column 3 presents average demographics of American adults.

Table 3: **Survey Response and Treatment Compliance Rates**

| Variable | (1) Treatment Mean/SD | (2) Control Mean/SD | T-test P-value (1)-(2) |
|---|---|---|---|
| Completed endline survey | 0.99 | 0.98 | 0.54 |
| | (0.11) | (0.12) | |
| Share of text messages completed | 0.92 | 0.93 | 0.45 |
| | (0.20) | (0.18) | |
| Completed post-endline survey | 0.95 | 0.92 | 0.07* |
| | (0.23) | (0.26) | |
| Share days deactivated | 0.90 | 0.02 | 0.00*** |
| | (0.29) | (0.13) | |
| N | 580 | 1081 | |

Notes: Columns 1 and 2 present survey response and treatment compliance rates for the Treatment and Control groups in the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. Column 3 presents p-values of tests of differences in response rates between the two groups.

Table 4: **Most Common Descriptions of Facebook Use Changes**

| Phrases Used More Often by Treatment | | | Phrases Used More Often by Control | | |
|---|---|---|---|---|---|
| Phrase | % Treatment | % Control | Phrase | % Treatment | % Control |
| not use facebook anymor | 0.91 | 0 | ha not chang | 6.73 | 17.05 |
| not spend much time | 1.09 | 0.38 | not chang sinc particip | 0 | 0.85 |
| spend less time facebook | 0.91 | 0.28 | ha not chang sinc | 0.18 | 1.42 |
| have not use facebook | 0.73 | 0.19 | chang sinc particip studi | 0 | 0.76 |
| not use facebook much | 0.73 | 0.19 | way use facebook ha | 0.18 | 1.33 |
| spend lot less time | 0.73 | 0.28 | awar much time spend | 0 | 0.66 |
| use much less | 2.91 | 0.66 | usag ha not chang | 0 | 0.66 |
| definit use facebook less | 0.55 | 0.19 | chang way use facebook | 0.18 | 1.23 |
| use facebook lot less | 0.55 | 0.19 | not chang | 7.27 | 19.03 |
| use facebook much less | 0.55 | 0.19 | ha not | 8.36 | 19.89 |
| not use facebook | 3.09 | 1.04 | more awar much time | 0.18 | 1.04 |
| use littl bit less | 0.55 | 0.28 | way use facebook | 0.55 | 2.75 |
| have not use | 1.27 | 0.19 | not think chang much | 0 | 0.47 |
| ha not chang use | 0.73 | 0.47 | not chang much use | 0 | 0.47 |
| use facebook anymor | 0.91 | 0.09 | use facebook slightli less | 0 | 0.47 |
| think use less | 1.64 | 0.47 | not much ha chang | 0 | 0.47 |
| no ha not chang | 0.55 | 0.38 | facebook ha not chang | 0.73 | 2.18 |
| use news app | 0.73 | 0.09 | much time spend | 0.18 | 1.52 |
| not use anymor | 0.73 | 0.09 | use facebook ha not | 0.55 | 1.70 |
| stop use facebook | 0.91 | 0.19 | use slightli less | 0 | 0.95 |

Notes: The post-endline survey included the following question with an open response text box: “How has the way you use Facebook changed, if at all, since participating in this study?” For all responses, we stemmed words, filtered out stop words, then constructed all phrases of length $l = \{1, 2, 3, 4\}$ words. For each phrase $p$ of length $l$, we calculated the number of occurrences of that phrase in Treatment and Control group responses ($f_{plT}$ and $f_{plC}$) and the number of occurrences of length-$l$ phrases that are *not* phrase $p$ in Treatment and Control responses ($f_{\sim plT}$ and $f_{\sim plC}$). We then constructed Pearson's $\chi^2$ statistic:

$$\chi^2 = \frac{\left(f_{plT} f_{\sim plC} - f_{plC} f_{\sim plT}\right)^2}{\left(f_{plT} + f_{plC}\right)\left(f_{plT} + f_{\sim plT}\right)\left(f_{plC} + f_{\sim plC}\right)\left(f_{\sim plT} + f_{\sim plC}\right)}.$$

This table presents the 20 phrases with the highest $\chi^2$ that were most commonly written by the Treatment and Control groups. The % Treatment and % Control columns present the share of people in the respective group whose responses included each phrase.

Table 5: **Estimating State Dependence and Misprediction**

| | Month 2 at midline $(-1 \times \phi)$ | | Month 2 at endline - at midline $(-1 \times \alpha)$ | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Share of time deactivated | 58.89*** | 92.06*** | -14.36*** | -18.22** |
| | (2.29) | (5.87) | (2.60) | (7.73) |
| Observations | 1,656 | 1,656 | 1,634 | 1,634 |
| Treatment mean month 2 WTA as of midline | 103 | 136 | 103 | 135 |
| Winsorized maximum WTA | 170 | 1,000 | 170 | 1,000 |

Notes: This table presents estimates of Equation (10). Standard errors are in parentheses. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Figure 1: **Experimental Design**

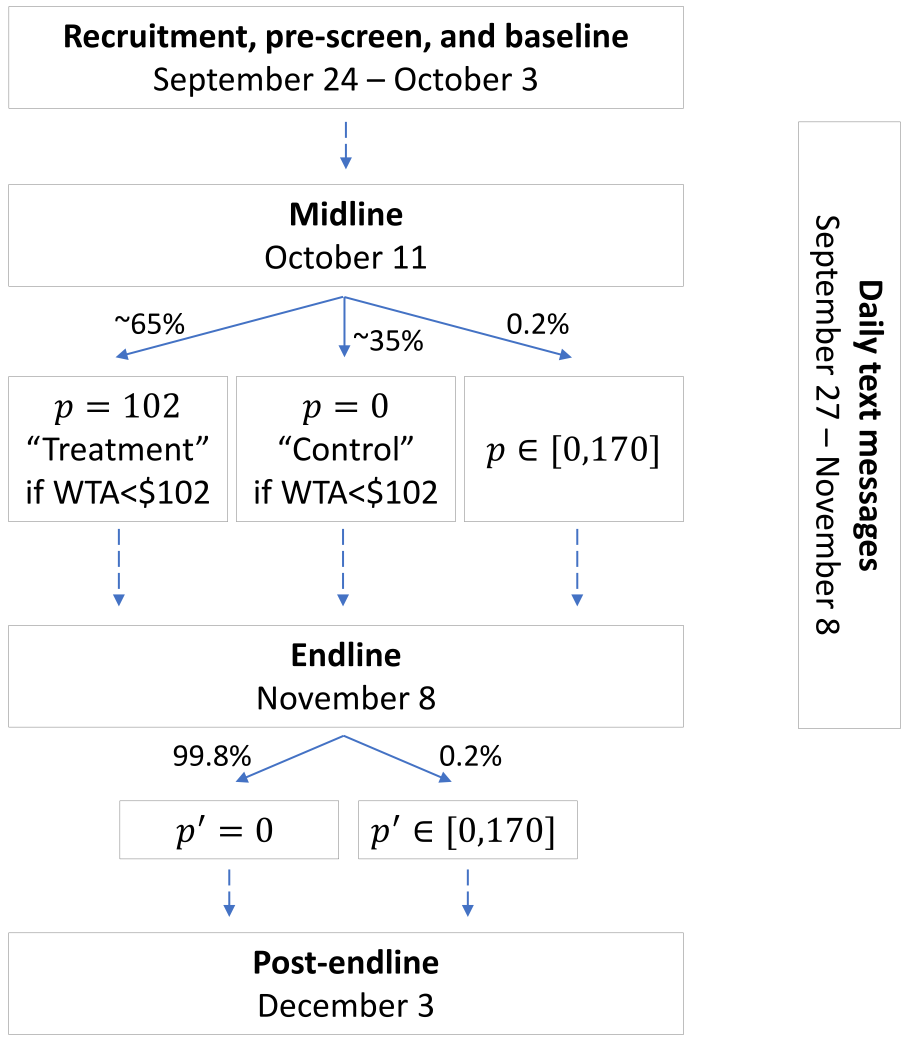
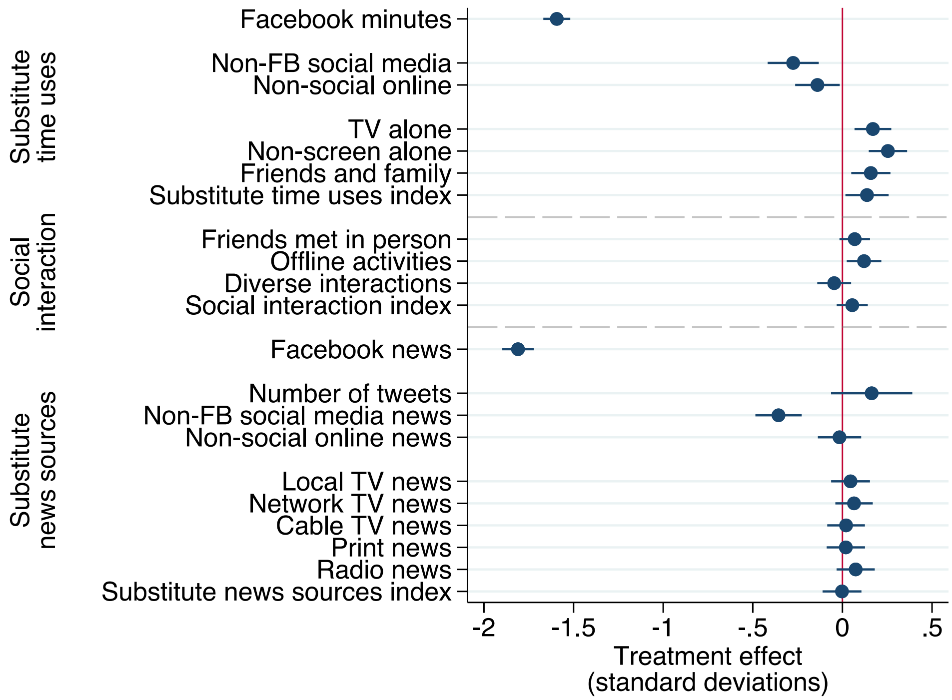
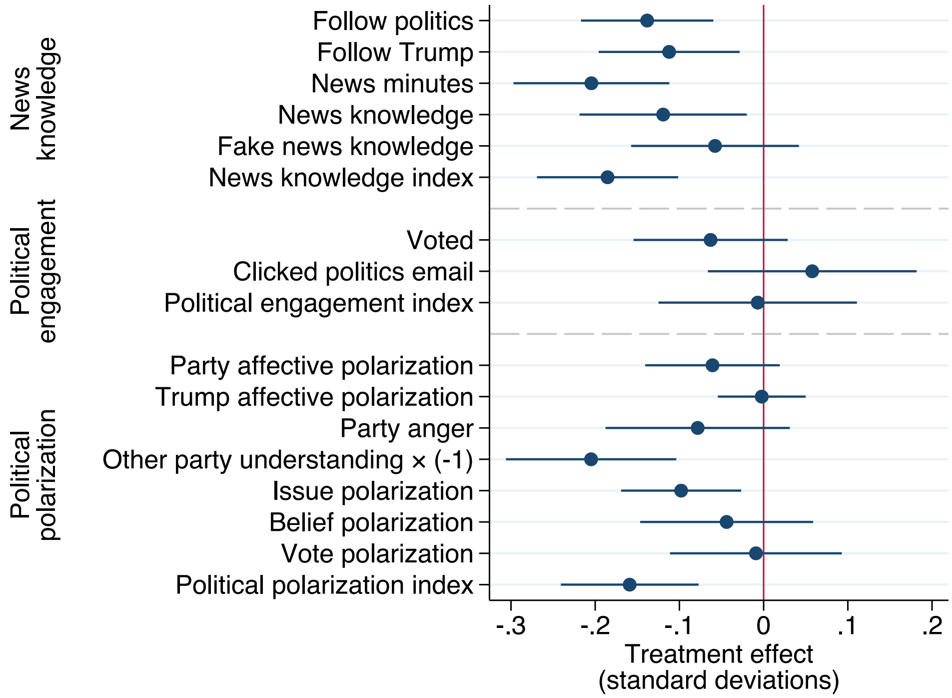


Figure 2: **Substitutes for Facebook**



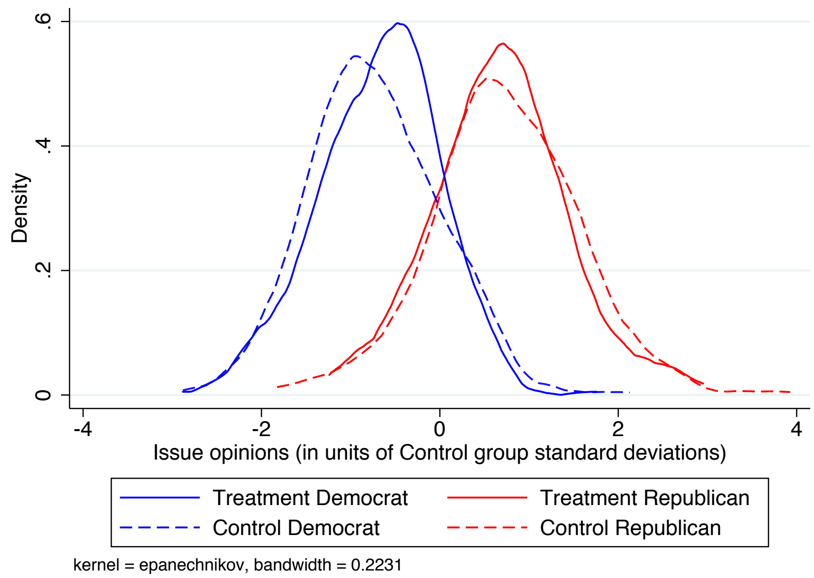

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions. Facebook minutes is not included in the substitute time uses index, and news from Facebook is not included in the substitute news sources index, so we visually separate these two variables from the other variables in their respective families. We also visually separate online and offline time uses and news sources, although all online and offline substitutes enter their respective indexes.

Figure 3: **Effects on News and Political Outcomes**



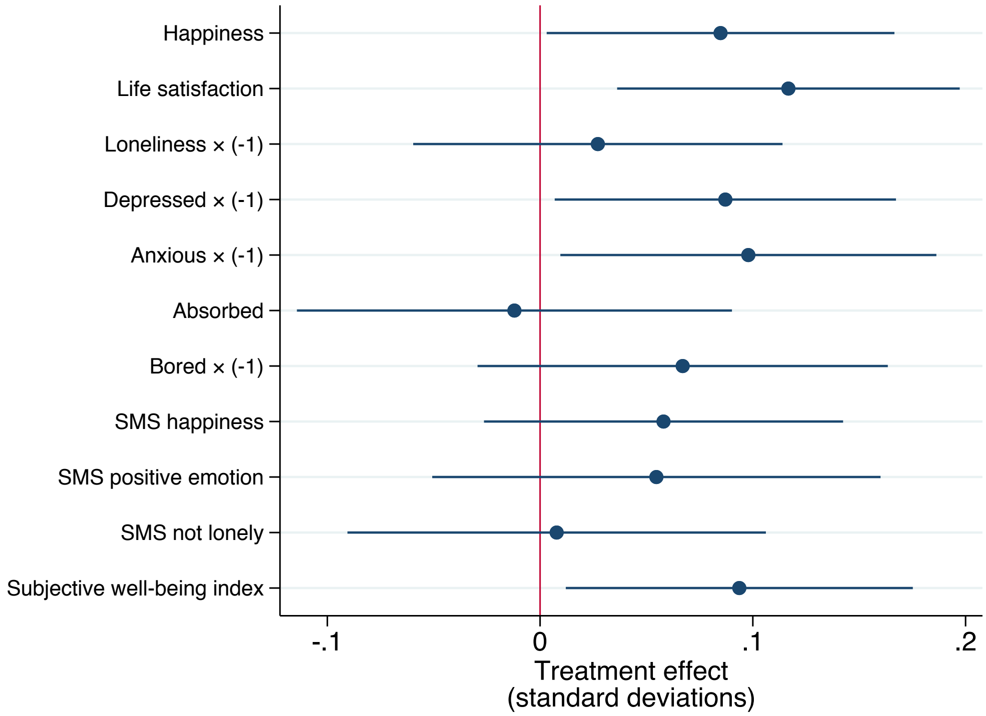

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure 4: **Issue Opinions by Party at Endline**



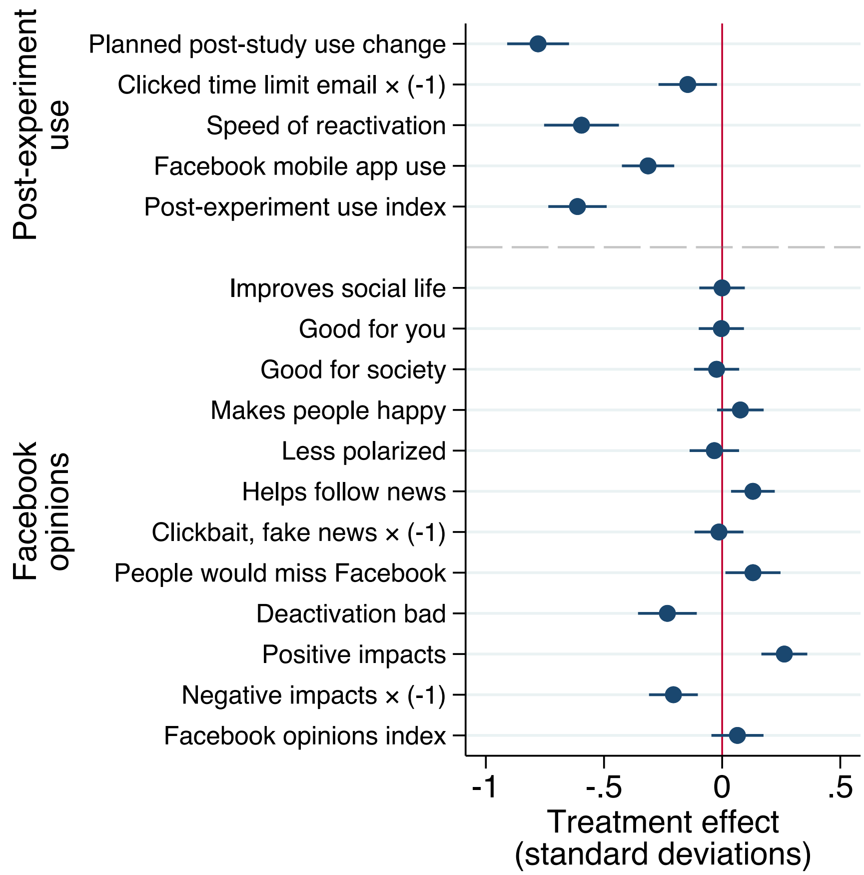

Notes: This figure presents kernel density plots of issue opinions for Democrats and Republicans in Treatment and Control at endline. Issue opinions are attitudes about nine current political issues on a scale from -5 to +5, such as "To what extent do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States." See Appendix B.1 for a list of all nine issue questions. To construct the *issue opinions* measure, for each issue question $q$, we normalize responses by the standard deviation in the Control group, determine Democrats' and Republicans' average responses $\mu_q^D$ and $\mu_q^R$, re-center so that $\mu_q^D + \mu_q^R = 0$, and re-sign so that $\mu^R > 0$. Define $\tilde{y}_{iq}$ as individual $i$'s normalized, re-centered, and re-signed response to question $q$. $\tilde{y}_{iq}$ thus reflects the strength of individual $i$'s agreement with the average Republican. Define $\sigma_q$ as the Control group within-person standard deviation of $\tilde{y}_{iq}$ for question $q$. This measures how much people's views change between baseline and endline, and allows us to place higher weight on issues about which views are malleable over the deactivation period. The preliminary *issue opinion* measure is $Y_i = \sum_q \tilde{y}_{iq}\sigma_q$, and the final *issue opinion* measure plotted in the figure is $Y_i$ divided by the Control group standard deviation.

Figure 5: **Effects on Subjective Well-Being**



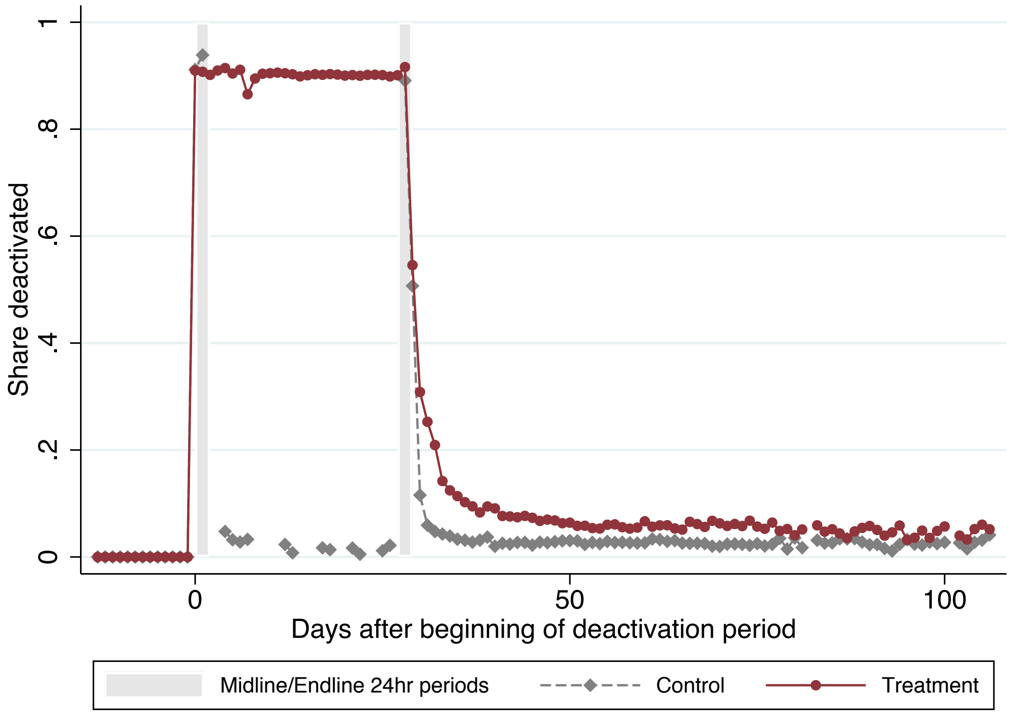

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure 6: **Effects on Post-Experiment Facebook Use and Opinions**



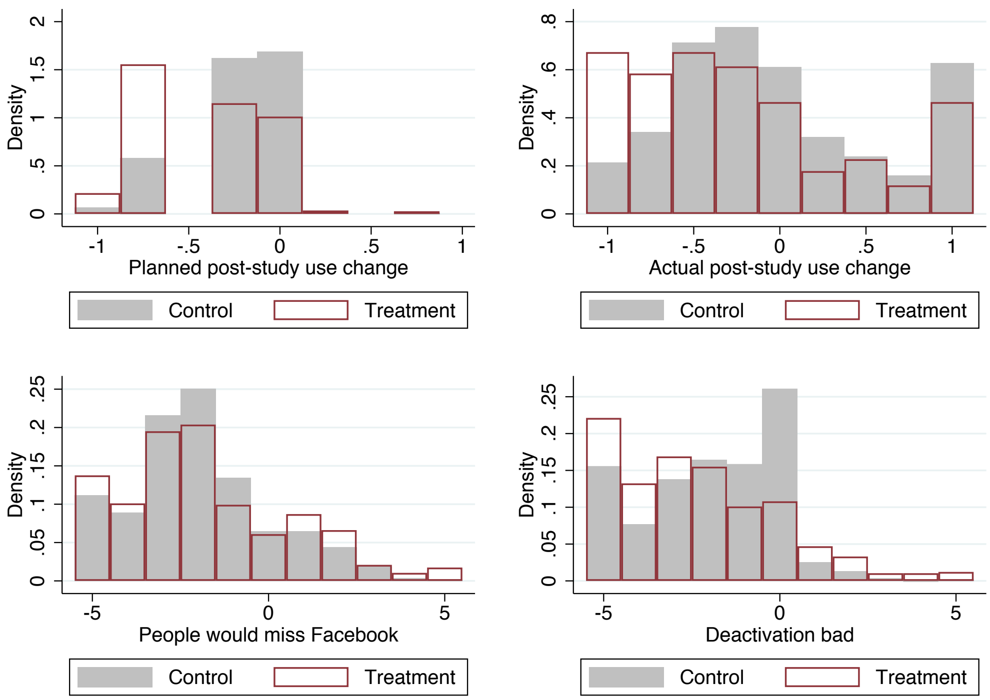

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure 7: **Probability of Being Deactivated**



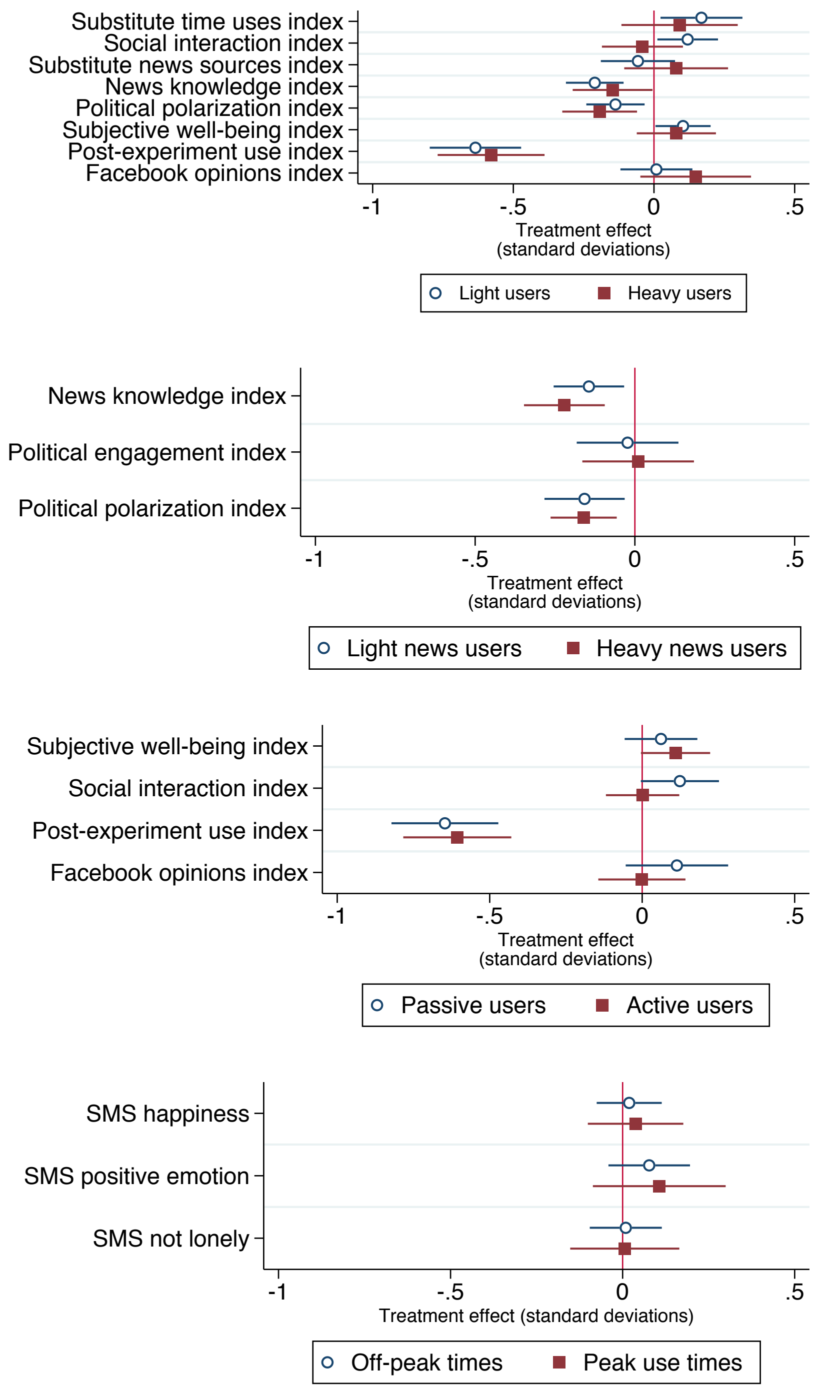

Notes: This figure shows the share of the Treatment and Control groups that had their Facebook accounts deactivated, by day of the experiment, for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. The vertical gray areas reflect the 24 hour periods after midline and endline during which both Treatment and Control were instructed to deactivate.

Figure 8: **Effects on Post-Experiment Facebook Use and Key Opinions**



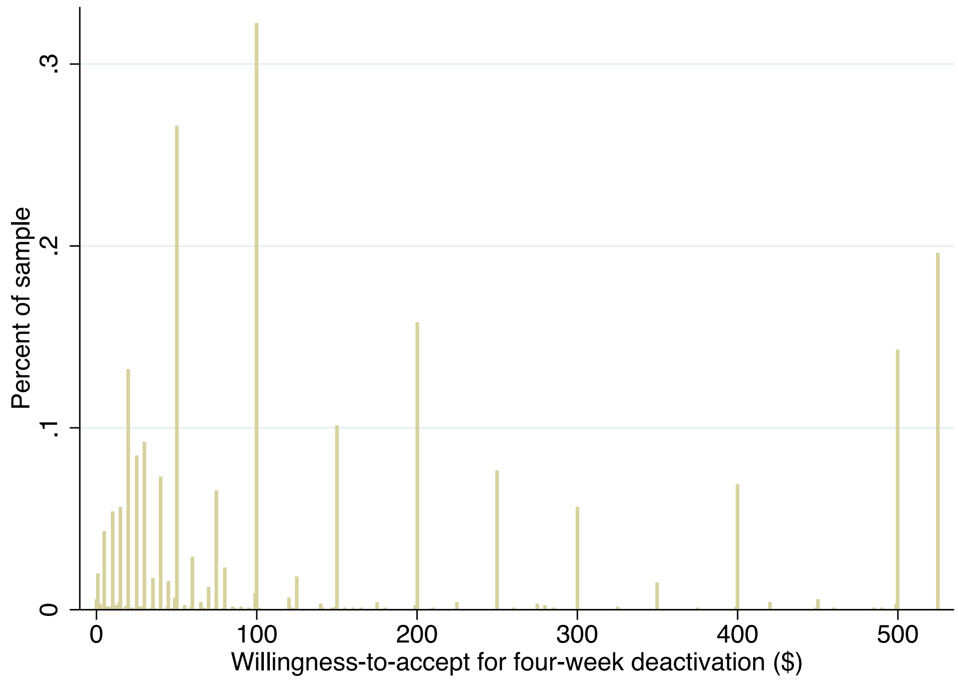

Notes: This figure presents the distribution of responses in Treatment and Control for planned and actual post-experiment Facebook use and two key measures of opinions about Facebook. See Section 2.3 for variable definitions.

Figure 9: **Heterogeneous Treatment Effects**



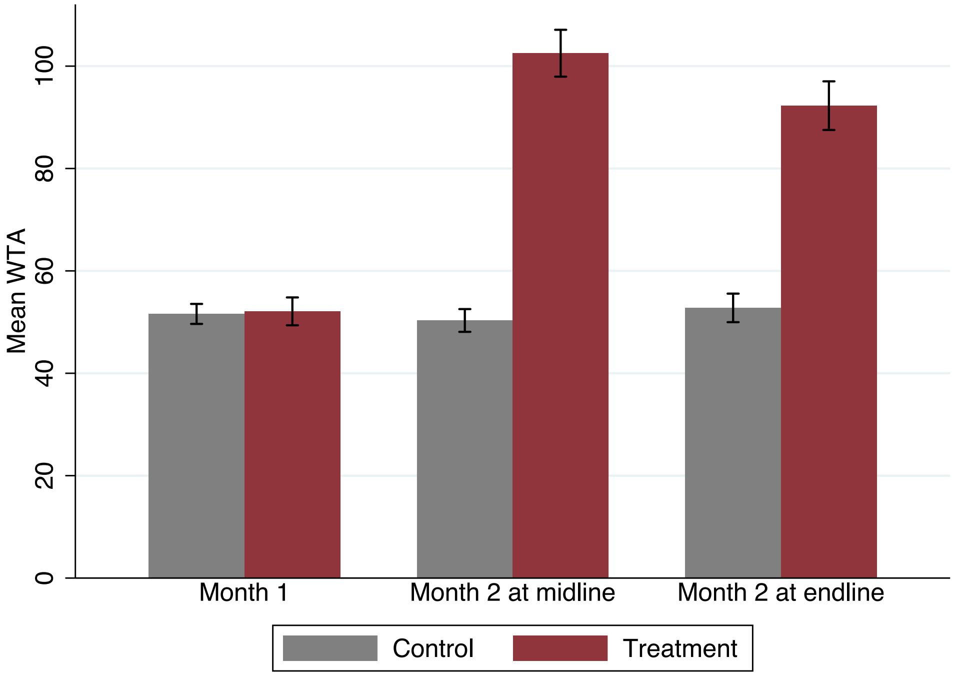

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1), for subgroups defined by the primary moderators in our pre-analysis plan. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure 10: **Distribution of Willingness-to-Accept to Deactivate Facebook After Midline**




Notes: This figure presents the distribution of willingness-to-accept to deactivate Facebook between midline and endline. All responses above $525 are plotted at $525.

Figure 11: **Average Valuation of Facebook in Treatment and Control**




Notes: This figure presents the mean willingness-to-accept (WTA) to deactivate Facebook in Treatment and Control, for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. The first pair of bars is the mean WTA for deactivation in "month 1," the four weeks after the midline survey. The second pair of bars is mean WTA for deactivation in "month 2," the four weeks after the endline survey, as elicited in the midline survey. The third pair of bars is mean WTA for deactivation in month 2, as elicited in the endline survey.

## Online Appendix: Not for Publication

The Welfare Effects of Social Media

*Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow*

Table A1: **Literature: Randomized Impact Evaluations of Facebook**

| Paper | N | Population | Intervention | Length | Enforcement | Outcomes | PAP |
|---|---|---|---|---|---|---|---|
| Gonzales and Hancock (2011) | 63 | College | Look at profile vs. mirror | 3 minutes | None | Self-esteem | No |
| Deters and Mehl (2012) | 86 | College | Post more status updates | 1 week | Scrape profile* | SWB | No |
| Mabe, Forney, and Keel (2014) | 84 | College women | Browse Facebook vs. research ocelots | 20 minutes | None | Eating disorder risk | No |
| Sagioglu and Greitemeyer (2014) | 263 | MTurk | Browse Facebook | 20 minutes | None | SWB | No |
| Fardouly and Vartanian (2015) | 112 | College women | Browse Facebook vs. other website | 10 minutes | None | Body image, mood | No |
| Verduyn et al. (2015) | 84 | College | Active vs. passive use | 10 minutes | Screen monitoring* | SWB | No |
| Theocharis and Lowe (2016) | 197 | Greek, without accounts | Sign up | 6 months | Payment sent to Facebook account* | Voting, civic engagement | No |
| Tromholt (2016) | 886** | Danish | Not log in | 1 week | Self-report | SWB | No |
| Marotta and Acquisti (2017) | 455 | MTurk | Block Facebook and YouTube during work hours | 2 weeks | Install blocking software | Work productivity | No |
| Hunt et al. (2018) | 111 | College | Limit social media to 10 minutes/day | 4 weeks | Weekly time use screen shots | SWB | No |
| Vanman, Baker, and Tobin (2018) | 123 | Australian | Not use Facebook | 5 days | None | Stress, SWB | No |
| Mosquera et al. (2018) | 151$^{\dagger}$ | College | Not log in | 1 week | Check "last active" | News, SWB, WTA*** | No |
| Allcott, Braghieri, Eichmeyer, and Gentzkow (2018) | 1,639 | US Facebook ads | Deactivate | 4 weeks | Check URLs | News, voting, polarization, SWB, WTA, WTA changes | Yes |

Notes: "N" is the number of people in the main empirical analysis, after attition. $^{\dagger}$1,765 people began this study, but 151 people were randomized and completed the endline survey. *Instead of analyzing as a randomized encouragement design, these studies dropped participants who did not comply with the treatment conditions. **This study had an 12 percent attrition rate in treatment and a 26 percent attrition rate in control. ***This study elicited WTA to participate in the experiment, which involved a 50 percent chance of Facebook deactivation plus completing a survey, and a 50 percent chance of only completing a survey.

# A Experimental Design Appendix

Figure A1: **Facebook Advertisement Used for Recruitment**

Figure A2: **Post-Endline Social Media Time Limit Email**

Stanford University

Dear ,

Thank you for participating in our study about Facebook use!

With the end of the study, we wanted to provide resources that can help you manage the technology in your life. We know how difficult it can be to stay away from or limit the time you spend on social media and other technologies.

There are tools that can help you track and limit your social media usage on your smartphone. If you would like to learn about them, click here if you have an iPhone and click here if you have an Android device. If you would like further information about other ways to curb smartphone use, here is a TIME magazine article we hope you find useful:

**Learn more about ways to limit smartphone use**

Thanks again for your participation in the study, and we wish you all the best.

Sarah, Luca, Kelly, Hunt, Matt, and Raj

**The Stanford Online Experience Research Study Research Team**

Figure A3: **Post-Endline Politics Email**

The 2018 midterm elections are over, but you can still participate and make your voice heard in important national, state, and local issues. There are many ways you might make a difference in future elections.

If you support Democratic candidates you might:

- Volunteer to help Democratic candidates in your community here.
- Sign a petition denouncing recent voter ID laws here.

**Donate to the Democratic Party to help prepare for future elections**

If you support Republican candidates you might:

- Volunteer to help Republican candidates in your community here.
- Sign a petition to encourage people to stand for the National Anthem here.

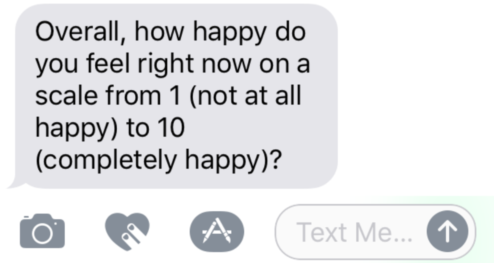
**Donate to the Republican Party to help prepare for future elections**

We hope you find these resources useful and engage with issues that matter to you.

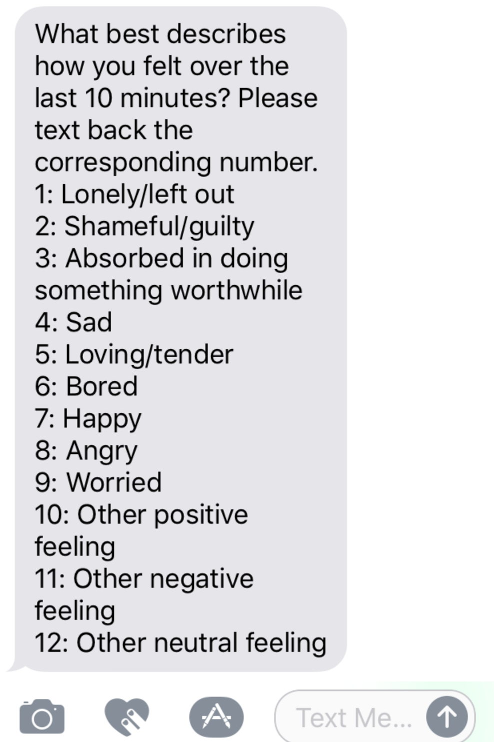
The Stanford Online Experience Research Study Research Team

Figure A4: **Subjective Well-Being Text Messages**

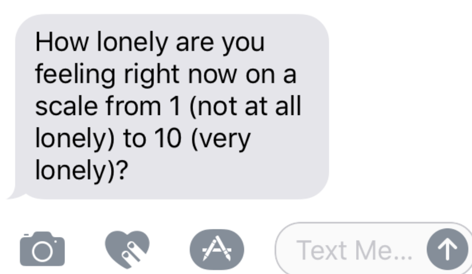
(a) **Happiness**




(b) **Primary Emotion**




(c) **Loneliness**



# B Variable Definitions and Descriptive Statistics

## B.1 Variable Definitions by Family

| Variable name | Question text |
|---|---|
| | *Substitute time uses* |
| Facebook minutes | On an average day in the past 4 weeks, how many minutes would you say you spent on Facebook, including through the Facebook app on your phone? *(not included in substitute time uses index)* |
| (At baseline) | On an average day in the last 4 weeks, how much free time (i.e. excluding work) did you spend... [0 minutes, Between 1 and 30 minutes, Between 31 minutes and 1 hour, Between 1 and 2 hours, Between 2 and 3 hours, More than 3 hours] |
| (At endline) | In the last 4 weeks, relative to what is typical for you, would you say you spent more or less of your free time (i.e. excluding work)... [A lot less, A little less, Same, A little more, A lot more] |
| Non-FB social media | ...using social media apps other than Facebook? |
| Non-social online | ...online (on your computer, tablet, smartphone, etc.) for things other than social media? |
| TV alone | ...watching TV or movies by yourself? |
| Non-screen alone | ...on non-screen activities (e.g. cooking, reading books, exercising – anything without an electronic screen in front of you) by yourself? |
| Friends and family | ...doing anything with friends and family (in person)? |
| | *Social interaction* |
| Friends met in person | List the first names of as many of the friends you met in person last week that you can think of in 1 minute (if none, enter "none"). Separate the names using commas (","). |
| Offline activities | Which of the following activities did you do at least once last week? Check all that apply<br>Go out for dinner<br>Go to the cinema<br>Talk to friends on the phone<br>Go to a party<br>Get together with friends<br>Go to a shopping mall<br>Spend time with your parents |

| | |
|---|---|
| | Spend time with your kids |
| Diverse interactions | Interact with someone who voted the opposite way as you in the last presidential election |
| | Interact with someone from another country |

| *Substitute news sources* | |
|---|---|
| (At baseline) | Over the past four weeks, how often did you... [Never, Hardly Ever, Sometimes, Fairly Often, Very Often] |
| (At endline) | In the last 4 weeks, relative to what is typical for you, would you say you spent more or less time... [A lot less, A little less, Same, A little more, A lot more] |
| Facebook news | ...get news from Facebook *(not included in substitute news sources index)* |
| Print news | ...read any newspapers in print? |
| Radio news | ...listen to the news on the radio? |
| Local TV news | ...watch local television news? |
| Network TV news | ...watch national evening network television news (such as ABC World News, CBS Evening News, or NBC Nightly News)? |
| Cable TV news | ...watch cable television news (such as CNN, the Fox News cable channel, or MSNBC)? |
| Non-FB social media news | ...get news from social media sites other than Facebook (e.g. Twitter or Snapchat)? |
| Non-social online news | ...get news from news websites or apps other than social media? |
| Number of tweets | ln(1+number of tweets in past four weeks) |

| *News knowledge* | |
|---|---|
| Follow politics | Thinking back over the last 4 weeks, how closely did you follow US politics? [Not at all closely, somewhat closely, rather closely, very closely] |
| Follow Trump | Thinking back over the last 4 weeks, how closely did you follow news about President Trump? [Not at all closely, somewhat closely, rather closely, very closely] |
| News minutes | On an average day of the last 4 weeks, how many minutes did you spend watching, reading or listening to the news (including news via social media)? [text box] |
| News knowledge | Of the following news events, which ones do you think are true, and which ones do you think are false? [True, False, Unsure] |

| | |
|---|---|
| (At baseline) *True statements* | Tension in trade negotiations escalated between the United States and China, with President Trump announcing tariffs on $200 billion worth of goods. |
| | An off-duty Dallas police officer entered the apartment of an African-American neighbor and shot and killed the unarmed neighbor. |
| | Deputy Attorney General Rod Rosenstein early in his tenure suggested secretly recording President Trump and recruiting cabinet members to remove him from office. |
| | The Trump administration set the maximum number of refugees that can enter the country in 2019 to 30,000. |
| | Michael Cohen, President Donald Trump's former personal attorney, agreed to cooperate with the Mueller investigation team and discuss Trump's business dealings with Russia. |
| | President Trump blasted Attorney General Jeff Sessions for the indictments of two lawmakers who supported Trump during the 2016 election. |
| | CBS chief executive Les Moonves resigned after multiple sexual misconduct allegations. |
| *False statements* | President Trump's former campaign chairman Paul Manafort refused deal to cooperate with the Mueller investigation team in exchange for legal charges against him being dropped. |
| | President Trump spoke at the funeral of former Arizona Senator John McCain, honoring the late McCain's wish. |
| | Hurricane Florence caused more than 300 deaths. |
| (At endline) *True statements* | A prominent Saudi Arabian journalist who was critical of the country's government was killed inside the Saudi Arabian consulate in Istanbul. |
| | In the weeks preceding the midterm elections, several high-profile Democrats, including Barack Obama and Hillary Clinton, were sent packages containing explosive devices. |
| | A mass shooting fueled by anti-Semitic sentiment took place in a synagogue in Pittsburgh. |
| | President Trump announced he plans to sign an executive order to prevent second-generation immigrants born in the United States from automatically being granted US citizenship. |
| | The Department of Justice charged a Russian national allegedly involved in a wide-ranging online disinformation campaign aimed at influencing the Midterm elections. |
| | One of the women who made allegations against Supreme Court Justice Brett Kavanaugh has admitted to investigators that the allegations were fabricated. |

| | |
|---|---|
| | Attorney General Jeff Sessions resigned at President Trump's request. |
| *False statements* | Harvard University recently stood trial for allegedly discriminating against African-American applicants in its admission process. |
| | Far-right candidate Jair Bolsonaro recently won an election to become the President of Argentina. |
| | Senator Elizabeth Warren's DNA test results show that she has no native American ancestry. |
| Fake news knowledge (At baseline) | After researcher Dr. Christine Blasey Ford accused Supreme Court nominee Brett Kavanaugh of sexual assault, it is revealed that Kavanaugh's mother once ruled against Dr. Blasey Ford's parents in a foreclosure case. |
| | CNN's Anderson Cooper reported deceptively on Hurricane Florence, standing in a ditch to create the misleading impression that he was filming amidst waist-deep floodwaters. |
| | Mayor Carmen Yulín Cruz of San Juan was arrested for misappropriating $3 million in disaster relief funds intended for the victims of Hurricane Maria in Puerto Rico. |
| | Clerk refused to sell gas to a man fleeing hurricane Florence over a Trump bumper sticker. |
| | WikiLeaks released an email showing that Hillary Clinton's presidential campaign bribed prominent Republicans to oppose Donald Trump during the 2016 election. |
| (At endline) | Billionaire George Soros was revealed to be one of the funders of a caravan of Central American emigrants traveling through Mexico to the US border. |
| | A Russian feminist activist poured bleach on men who were "manspreading" on the train ("manspreading" refers to men sitting in public transport with legs wide apart, thereby covering more than one seat). |
| | In a recent vote, all Democrats in Congress voted against a 2.8% cost of living allowance in Social Security benefits. |
| | Cesar Sayoc, suspect in an act of domestic terrorism directed at vocal critics of President Trump, was a registered Democrat. |
| | None of the 154 mass shootings in 2018 was committed by a black man, illegal alien, or woman. |
| | *Political engagement* |
| Voted | Takes value 1 if recorded as having voted in 2018 midterm, and 0 otherwise |

| | |
|---|---|
| Clicked politics email | Takes value 1 if clicked on any link in the post-endline politics email, and 0 otherwise |
| *Political polarization* | |
| Party affective polarization | Thinking back over the last 4 weeks, how warm or cold did you feel towards the parties and the president on the feeling thermometer? |
| Trump affective polarization | Thinking back over the last 4 weeks, how warm or cold did you feel towards the parties and the president on the feeling thermometer? |
| Party anger | List as many recent (last 4 weeks) news events you can think of that made you angry at the [Republican/Democratic] Party. (If more than 5, just list those 5 that left you most angry. If less than 5, list less. If none, enter "none" in the first textbox.) |
| Other party understanding × (-1) | Thinking back over the last 4 weeks, how often did you see news that made you better understand the point of view of the Republican Party? [Never, Once, Two or three times, Four times or more]<br>Thinking back over the last 4 weeks, how often did you see news that made you better understand the point of view of the Democratic Party? [Never, Once, Two or three times, Four times or more] |
| Issue polarization | To what extent do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States? (Pew Research Center 2018b)<br>Overall, would you say that blacks or whites are treated more fairly in dealing with the police? (Pew Research Center 2016)<br>Do you think that employers firing men who have been accused of sexual harassment or assault before finding out all the facts is a major or a minor problem? (Pew Research Center 2018d)<br>As you may know, Brett Kavanaugh is a federal judge who has been nominated to serve on the Supreme Court. Would you like to see the Senate vote in favor of Kavanaugh serving on the Supreme Court, or not? (Gallup 2018b)<br>On the whole, do you think immigration is a good thing or a bad thing for this country today? (Pew Research Center 2018e)<br>How confident, if at all, are you that the Justice Department special counsel Robert Mueller will conduct a fair investigation into Russian involvement in the 2016 election? (Pew Research Center 2018c)<br>In general, do you feel that the laws covering the sale of firearms should be made less strict, more strict, or kept as they are now? (Gallup 2018c) |

| | |
|---|---|
| | In presenting the news dealing with political and social issues, do you think that news organizations deal fairly with all sides, or do they tend to favor one side? (Pew Research Center 2017) |
| | To what extent do you think President Trump is honest and trustworthy? (Gallup 2018a) |
| Belief polarization | Level of agreement with co-partisans on beliefs questions |
| Vote polarization | Strength of generic ballot preference for co-partisan candidate (see Voted Republican question) |
| | *Subjective well-being* |
| Happiness | Over the last 4 weeks, I think I was [1 (not a very happy person) ... 7 (a very happy person)] |
| | Over the last 4 weeks, compared to most of my peers, I think I was [1 (less happy) ... 7 (more happy)] |
| Life satisfaction | Below are three statements that you may agree or disagree with. Indicate your agreement with each item and please be open and honest in your responding. [Strongly disagree, Disagree, Slightly disagree, Neither agree nor disagree, Slightly agree, Agree, Strongly agree] |
| | In most ways my life during the past 4 weeks was close to ideal. |
| | The conditions of my life during the past 4 weeks were excellent. |
| | During the past 4 weeks, I was satisfied with my life. |
| Loneliness × (-1) | How often did you feel that you lacked companionship over the past four weeks [Hardly ever, Some of the time, Often] |
| | How often did you feel left out over the past four weeks [Hardly ever, Some of the time, Often] |
| | How often did you feel isolated from others over the past four weeks [Hardly ever, Some of the time, Often] |
| | Below are some ways you might have felt or behaved in the past 4 weeks. Please tell us how much of the time during the past 4 weeks: [1 None or almost none of the time, 2, 3, 4 All or almost all of the time] |
| Depressed × (-1) | ... you felt depressed. |
| Anxious × (-1) | ... you felt anxious. |
| Absorbed | ... you were absorbed in doing something worthwhile. |
| Bored × (-1) | ... you felt bored. |
| SMS happiness | Overall, how happy do you feel right now on a scale from 1 (not at all happy) to 10 (completely happy)? |

| | |
|---|---|
| SMS positive emotion | What best describes how you felt over the last 10 minutes? Please text back the corresponding number. [1: Lonely/left out 2: Shameful/guilty 3: Absorbed in doing something worthwhile 4: Sad 5: Loving/tender 6: Bored 7: Happy 8: Angry 9: Worried 10: Other positive feeling 11: Other negative feeling 12: Other neutral feeling |
| SMS not lonely | How lonely are you feeling right now on a scale from 1 (not at all lonely) to 10 (very lonely)? |

| *Post-experiment use* | |
|---|---|
| Planned post-study use change | After going through this study, how much more or less time do you plan to spend on Facebook compared to before you started the study? |
| Clicked time limit email × (-1) | Takes value 1 if clicked on any link in the post-endline social media time limit email, and 0 otherwise |
| Speed of reactivation | (-1) × ln(1+number of days deactivated after 24-hour post-endline deactivation period) |
| Facebook mobile app use | [if have an iPhone] Please write down the amount of screen time you spent on the Facebook app according to your battery report.<br>[if do not have an iPhone] How many hours would you say you spent on the Facebook app on your phone in the past seven days, in total? |

| *Facebook opinions* | |
|---|---|
| Improves social life | To what extent do you think Facebook improves or worsens people's social lives? |
| Good for you | To what extent do you think Facebook is good or bad for you? |
| Good for society | To what extent do you think Facebook is good or bad for society? |
| Makes people happy | To what extent do you think using Facebook makes people more or less happy? |
| People would miss Facebook | To what extent do you agree or disagree with the following statement: "If people spent less time on Facebook, they would soon realize that they don't miss it."? *(We multiply responses by -1, so more agreement with the statement is more negative.)* |
| Helps follow news | To what extent do you think Facebook helps people follow the news better? |
| Clickbait, fake news × (-1) | To what extent do you think Facebook exposes people to clickbait or false news stories? |
| Less polarized | To what extent do you think Facebook makes people more or less politically polarized? |

| | |
|---|---|
| Deactivation bad | As part of this study, you were asked to deactivate your Facebook account for [24 hours/4 weeks]. To what extent do you think that deactivating your account was good or bad for you? *(We multiply responses by -1, so responding that deactivation was good is more negative.)* |
| Positive impacts | What are the most important positive impact(s) that Facebook has on your life? [text box] |
| Negative impacts × (-1) | What are the most important negative impact(s) that Facebook has on your life? [text box] |
| | *Secondary outcomes* |
| Voted Republican | If the elections for US Congress were being held today, would you vote for the Republican Party's candidate or the Democratic Party's candidate for Congress in your district? [Republican candidate, Democratic candidate, Other/don't know]<br>[If would vote for Republican or Democratic candidate] How convinced are you about whether to vote for the Republican candidate or the Democratic candidate? [slider from 0 to 100] |
| Voted (self-report) | Did you [midline: Do you plan to] vote in the midterm elections on November 6th, 2018? |
| Contributions | ln(1+FEC contributions between October 12 and November 10) |
| | *Moderators* |
| Time of day | At what times of day do you usually use Facebook the most? [Morning (6AM-12 noon), Afternoon (12 noon-5PM), Evening (5-9PM), Night (9PM-midnight), Late night/early morning (midnight-6AM) |
| Active browsing | People talk about two different ways to use Facebook:<br>"Active" users often post status updates, comment on other people's walls and pictures, post photos, etc.<br>"Passive" users mostly check out their news feeds and/or other people's photos and profiles but don't comment or interact much with others on the site.<br>Which would you say describes your Facebook use best?<br>What share of your time on Facebook do you spend interacting one-on-one with people you care about (for example, commenting on their posts or sending them private messages)? |
| Get news from Facebook | Over the past four weeks, how often did you ... get news from Facebook [Never, Hardly Ever, Sometimes, Fairly Often, Very Often] |

| | |
|---|---|
| Facebook minutes | On an average day in the past 4 weeks, how many minutes would you say you spent on Facebook, including through the Facebook app on your phone? |

## B.2 Descriptive Statistics

Table A3: **Descriptive Statistics: Substitutes for Facebook and News and Political Outcomes**

| | Mean | Standard deviation | Minimum value | Maximum value | N in regression |
|---|---|---|---|---|---|
| Facebook minutes | 59.53 | 37.38 | 0 | 120 | 1,639 |
| Non-FB social media | 2.97 | 0.93 | 1 | 5 | 1,639 |
| Non-social online | 3.28 | 0.88 | 1 | 5 | 1,639 |
| TV alone | 3.10 | 1.02 | 1 | 5 | 1,639 |
| Non-screen alone | 3.23 | 0.92 | 1 | 5 | 1,639 |
| Friends and family | 3.24 | 0.91 | 1 | 5 | 1,639 |
| Friends met in person | 1.36 | 0.85 | 0 | 3 | 1,639 |
| Offline activities | 3.06 | 1.53 | 0 | 8 | 1,639 |
| Diverse interactions | 0.99 | 0.79 | 0 | 2 | 1,639 |
| Facebook news | 2.98 | 1.05 | 1 | 5 | 1,639 |
| Number of tweets | 1.80 | 1.48 | 0 | 6 | 296 |
| Non-FB social media news | 3.04 | 1.03 | 1 | 5 | 1,639 |
| Non-social online news | 3.40 | 1.01 | 1 | 5 | 1,639 |
| Local TV news | 3.00 | 0.95 | 1 | 5 | 1,639 |
| Network TV news | 2.93 | 0.98 | 1 | 5 | 1,639 |
| Cable TV news | 2.93 | 1.01 | 1 | 5 | 1,639 |
| Print news | 2.72 | 0.95 | 1 | 5 | 1,639 |
| Radio news | 2.86 | 1.00 | 1 | 5 | 1,639 |
| Follow politics | 2.32 | 0.98 | 1 | 4 | 1,639 |
| Follow Trump | 2.09 | 0.92 | 1 | 4 | 1,639 |
| News minutes | 52.10 | 38.72 | 0 | 120 | 1,639 |
| News knowledge | 7.26 | 1.19 | 3 | 10 | 1,639 |
| Fake news knowledge | 2.72 | 0.74 | 0 | 5 | 1,639 |
| Voted | 0.77 | 0.42 | 0 | 1 | 1,639 |
| Clicked politics email | 0.02 | 0.15 | 0 | 1 | 1,651 |
| Party affective polarization | 53.21 | 34.37 | -86 | 100 | 1,455 |
| Trump affective polarization | 32.73 | 26.72 | -50 | 50 | 1,455 |
| Party anger | 1.97 | 2.27 | -5 | 6 | 1,450 |
| Other party understanding × (-1) | 1.00 | 1.54 | -4 | 4 | 1,450 |
| Issue polarization | 2.89 | 2.97 | -8 | 15 | 1,450 |
| Belief polarization | 2.16 | 5.21 | -15 | 17 | 1,450 |
| Vote polarization | 0.63 | 0.48 | -1 | 1 | 1,450 |

Notes: This table presents descriptive statistics for the dependent variables used in Equations (1) and (2). Survey outcomes were recorded in the endline or post-endline surveys. The mean, standard deviation, minimum, and maximum are for the prepared variables as used in the regressions, before normalizing to standard deviation of one, for the Control group: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p =$ \$0 to do so. See Section 2.3 for variable definitions. *Facebook minutes* and *news minutes* are winsorized at 120. *Number of tweets* is the natural log of one plus the number of tweets.

Table A4: **Descriptive Statistics: Subjective Well-Being, Post-Experiment Facebook Use and Opinions, and Secondary Outcomes**

| | Mean | Standard deviation | Minimum value | Maximum value | N in regression |
|---|---|---|---|---|---|
| Happiness | 4.47 | 1.41 | 1 | 7 | 1,639 |
| Life satisfaction | 12.26 | 4.78 | 3 | 21 | 1,639 |
| Loneliness × (-1) | -5.19 | 1.89 | -9 | -3 | 1,639 |
| Depressed × (-1) | 2.99 | 0.97 | 1 | 4 | 1,639 |
| Anxious × (-1) | 2.60 | 0.94 | 1 | 4 | 1,639 |
| Absorbed | 2.82 | 0.80 | 1 | 4 | 1,639 |
| Bored × (-1) | 2.93 | 0.88 | 1 | 4 | 1,639 |
| SMS happiness | 6.48 | 1.52 | 1 | 10 | 1,603 |
| SMS positive emotion | 0.53 | 0.25 | 0 | 1 | 1,606 |
| SMS not lonely | 7.60 | 1.70 | 1 | 10 | 1,604 |
| Planned post-study use change | -0.22 | 0.28 | -1 | 1 | 1,637 |
| Clicked time limit email × (-1) | -0.09 | 0.28 | -1 | 0 | 1,660 |
| Speed of reactivation | -0.41 | 0.69 | -4 | 0 | 1,661 |
| Facebook mobile app use | 52.80 | 38.76 | 0 | 120 | 1,219 |
| Improves social life | -0.39 | 1.93 | -5 | 5 | 1,639 |
| Good for you | -0.28 | 1.76 | -5 | 5 | 1,639 |
| Good for society | -0.53 | 1.86 | -5 | 5 | 1,639 |
| Makes people happy | -0.82 | 1.81 | -5 | 5 | 1,639 |
| Less polarized | -2.48 | 1.76 | -5 | 5 | 1,639 |
| Helps follow news | 0.31 | 2.41 | -5 | 5 | 1,639 |
| Clickbait, fake news × (-1) | -2.71 | 2.06 | -5 | 5 | 1,639 |
| People would miss Facebook | -1.97 | 1.99 | -5 | 5 | 1,639 |
| Deactivation bad | -1.91 | 1.93 | -5 | 5 | 1,639 |
| Positive impacts | 3.72 | 0.83 | 0 | 8 | 1,639 |
| Negative impacts × (-1) | -3.44 | 1.03 | -7 | 0 | 1,639 |
| Voted Republican | -0.36 | 0.68 | -1 | 1 | 1,639 |

Notes: This table presents descriptive statistics for the dependent variables used in Equations (1) and (2). Survey outcomes were recorded in the endline or post-endline surveys. The mean, standard deviation, minimum, and maximum are for the prepared variables as used in the regressions, before normalizing to standard deviation of one, for the Control group: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p =$ \$0 to do so. See Section 2.3 for variable definitions. *Facebook mobile app use* is winsorized at 120. *Positive impacts* and *negative impacts* are the natural log of one plus number of characters the participant wrote in the text box. Speed of reactivation is negative one times the natural log of one plus the number of days that the participant remained deactivated after 24-hour post-endline deactivation period), top-coded at the last day of measurement. *Contributions* is the natural log of one plus the dollar amount of FEC contributions made between October 12 and November 10, 2018.

Table A5: **Descriptive Statistics: Pre-Experiment Time Use**

| | Mean | Standard deviation | Minimum value | Maximum value |
|---|---|---|---|---|
| Facebook minutes | 74.5 | 35.5 | 20 | 120 |
| News minutes | 53.0 | 37.9 | 0 | 120 |
| Non-FB social media | 75.7 | 76.3 | 0 | 240 |
| Non-social online | 135.9 | 83.7 | 0 | 240 |
| TV alone | 95.5 | 82.8 | 0 | 240 |
| Non-screen alone | 105.9 | 79.2 | 0 | 240 |
| Friends and family | 130.4 | 83.4 | 0 | 240 |
| Facebook mobile app use | 60.0 | 38.9 | 0 | 120 |

Notes: This table presents descriptive statistics for pre-experiment time use, for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. These survey outcomes were recorded in the baseline and midline surveys. See Section 2.3 for variable definitions. *Facebook minutes*, *news minutes,* and *Facebook mobile app use* are winsorized at 120.

# C Tables of Treatment Effect Estimates

Table A6: **Treatment Effects: Substitutes for Facebook and News and Political Outcomes**

| | Treatment effect (original units) | Standard error (original units) | Treatment effect (SD units) | Standard error (SD units) | P-value | Sharpened FDR-adjusted q-value |
|---|---|---|---|---|---|---|
| Facebook minutes | -59.58 | 1.43 | -1.59 | 0.04 | 0.00 | 0.00 |
| Non-FB social media | -0.25 | 0.07 | -0.27 | 0.07 | 0.00 | 0.00 |
| Non-social online | -0.12 | 0.06 | -0.14 | 0.06 | 0.03 | 0.04 |
| TV alone | 0.17 | 0.05 | 0.17 | 0.05 | 0.00 | 0.00 |
| Non-screen alone | 0.23 | 0.05 | 0.25 | 0.05 | 0.00 | 0.00 |
| Friends and family | 0.14 | 0.05 | 0.16 | 0.06 | 0.00 | 0.01 |
| Friends met in person | 0.06 | 0.04 | 0.07 | 0.04 | 0.11 | 0.13 |
| Offline activities | 0.18 | 0.08 | 0.12 | 0.05 | 0.01 | 0.03 |
| Diverse interactions | -0.04 | 0.04 | -0.05 | 0.05 | 0.34 | 0.29 |
| Facebook news | -1.90 | 0.05 | -1.81 | 0.04 | 0.00 | 0.00 |
| Number of tweets | 0.24 | 0.17 | 0.16 | 0.12 | 0.16 | 0.16 |
| Non-FB social media news | -0.37 | 0.07 | -0.36 | 0.07 | 0.00 | 0.00 |
| Non-social online news | -0.02 | 0.06 | -0.02 | 0.06 | 0.79 | 0.44 |
| Local TV news | 0.04 | 0.05 | 0.05 | 0.06 | 0.41 | 0.33 |
| Network TV news | 0.06 | 0.05 | 0.06 | 0.05 | 0.22 | 0.20 |
| Cable TV news | 0.02 | 0.05 | 0.02 | 0.05 | 0.70 | 0.40 |
| Print news | 0.02 | 0.05 | 0.02 | 0.05 | 0.73 | 0.41 |
| Radio news | 0.07 | 0.05 | 0.07 | 0.05 | 0.17 | 0.16 |
| Follow politics | -0.14 | 0.04 | -0.14 | 0.04 | 0.00 | 0.00 |
| Follow Trump | -0.10 | 0.04 | -0.11 | 0.04 | 0.01 | 0.02 |
| News minutes | -7.92 | 1.83 | -0.20 | 0.05 | 0.00 | 0.00 |
| News knowledge | -0.14 | 0.06 | -0.12 | 0.05 | 0.02 | 0.03 |
| Fake news knowledge | -0.04 | 0.04 | -0.06 | 0.05 | 0.26 | 0.23 |
| Voted | -0.03 | 0.02 | -0.06 | 0.05 | 0.18 | 0.16 |
| Clicked politics email | 0.01 | 0.01 | 0.06 | 0.06 | 0.36 | 0.31 |
| Party affective polarization | -2.09 | 1.40 | -0.06 | 0.04 | 0.14 | 0.15 |
| Trump affective polarization | -0.06 | 0.71 | -0.00 | 0.03 | 0.93 | 0.50 |
| Party anger | -0.18 | 0.13 | -0.08 | 0.06 | 0.16 | 0.16 |
| Other party understanding × (-1) | -0.32 | 0.08 | -0.20 | 0.05 | 0.00 | 0.00 |
| Issue polarization | -0.29 | 0.11 | -0.10 | 0.04 | 0.01 | 0.02 |
| Belief polarization | -0.23 | 0.27 | -0.04 | 0.05 | 0.40 | 0.33 |
| Vote polarization | -0.00 | 0.02 | -0.01 | 0.05 | 0.86 | 0.47 |

Notes: This table presents local average treatment effects of Facebook deactivation estimated using Equation (1). Column 1 and Column 2 present the effect and standard error on un-normalized outcomes. Columns 3 and 4 present the effect and standard error on normalized outcomes, where outcomes are normalized so that the Control group endline distribution has a standard deviation of one. Columns 5 and 6 present the unadjusted p-value and sharpened False Discovery Rate-adjusted two-stage q-value, respectively.

Table A7: **Treatment Effects: Subjective Well-Being, Post-Experiment Facebook Use and Opinions, and Secondary Outcomes**

| | Treatment effect (original units) | Standard error (original units) | Treatment effect (SD units) | Standard error (SD units) | P-value | Sharpened FDR-adjusted q-value |
|---|---|---|---|---|---|---|
| Happiness | 0.12 | 0.06 | 0.08 | 0.04 | 0.04 | 0.06 |
| Life satisfaction | 0.56 | 0.20 | 0.12 | 0.04 | 0.00 | 0.01 |
| Loneliness × (-1) | 0.05 | 0.08 | 0.03 | 0.04 | 0.54 | 0.37 |
| Depressed × (-1) | 0.08 | 0.04 | 0.09 | 0.04 | 0.03 | 0.05 |
| Anxious × (-1) | 0.09 | 0.04 | 0.10 | 0.05 | 0.03 | 0.05 |
| Absorbed | -0.01 | 0.04 | -0.01 | 0.05 | 0.82 | 0.45 |
| Bored × (-1) | 0.06 | 0.04 | 0.07 | 0.05 | 0.17 | 0.16 |
| SMS happiness | 0.09 | 0.07 | 0.06 | 0.04 | 0.18 | 0.16 |
| SMS positive emotion | 0.01 | 0.01 | 0.05 | 0.05 | 0.31 | 0.28 |
| SMS not lonely | 0.01 | 0.09 | 0.01 | 0.05 | 0.88 | 0.47 |
| Planned post-study use change | -0.21 | 0.02 | -0.78 | 0.07 | 0.00 | 0.00 |
| Clicked time limit email × (-1) | -0.04 | 0.02 | -0.15 | 0.06 | 0.02 | 0.04 |
| Speed of reactivation | -0.41 | 0.06 | -0.60 | 0.08 | 0.00 | 0.00 |
| Facebook mobile app use | -12.15 | 2.19 | -0.31 | 0.06 | 0.00 | 0.00 |
| Improves social life | -0.00 | 0.09 | -0.00 | 0.05 | 0.99 | 0.50 |
| Good for you | -0.01 | 0.09 | -0.00 | 0.05 | 0.95 | 0.50 |
| Good for society | -0.04 | 0.09 | -0.02 | 0.05 | 0.63 | 0.37 |
| Makes people happy | 0.14 | 0.09 | 0.08 | 0.05 | 0.13 | 0.14 |
| Less polarized | -0.06 | 0.09 | -0.03 | 0.05 | 0.54 | 0.37 |
| Helps follow news | 0.31 | 0.11 | 0.13 | 0.05 | 0.01 | 0.01 |
| Clickbait, fake news × (-1) | -0.03 | 0.11 | -0.01 | 0.05 | 0.80 | 0.44 |
| People would miss Facebook | 0.26 | 0.12 | 0.13 | 0.06 | 0.03 | 0.04 |
| Deactivation bad | -0.45 | 0.12 | -0.23 | 0.06 | 0.00 | 0.00 |
| Positive impacts | 0.22 | 0.04 | 0.26 | 0.05 | 0.00 | 0.00 |
| Negative impacts × (-1) | -0.21 | 0.05 | -0.21 | 0.05 | 0.00 | 0.00 |
| Voted Republican | -0.04 | 0.02 | -0.07 | 0.04 | 0.06 | 0.08 |

Notes: This table presents local average treatment effects of Facebook deactivation estimated using Equation (1). Column 1 and Column 2 present the effect and standard error on un-normalized outcomes. Columns 3 and 4 present the effect and standard error on normalized outcomes, where outcomes are normalized so that the Control group endline distribution has a standard deviation of one. Columns 5 and 6 present the unadjusted p-value and sharpened False Discovery Rate-adjusted two-stage q-value, respectively.

Table A8: **Treatment Effects: Indices**

| | Treatment effect | Standard error | P-value | Sharpened FDR-adjusted q-value |
|---|---|---|---|---|
| Substitute time uses index | 0.14 | 0.06 | 0.03 | 0.03 |
| Social interaction index | 0.05 | 0.04 | 0.22 | 0.17 |
| Substitute news sources index | -0.00 | 0.06 | 0.97 | 0.47 |
| News knowledge index | -0.19 | 0.04 | 0.00 | 0.00 |
| Political engagement index | -0.01 | 0.06 | 0.91 | 0.47 |
| Political polarization index | -0.16 | 0.04 | 0.00 | 0.00 |
| Subjective well-being index | 0.09 | 0.04 | 0.02 | 0.03 |
| Post-experiment use index | -0.61 | 0.06 | 0.00 | 0.00 |
| Facebook opinions index | 0.06 | 0.06 | 0.25 | 0.17 |

Notes: This table presents local average treatment effects of Facebook deactivation on index outcomes estimated using Equation (1). Columns 1 and 2 present the effect and standard error, with indices normalized so that the Control group endline distribution has a standard deviation of one. Columns 3 and 4 present the unadjusted p-value and sharpened False Discovery Rate-adjusted two-stage q-value, respectively.

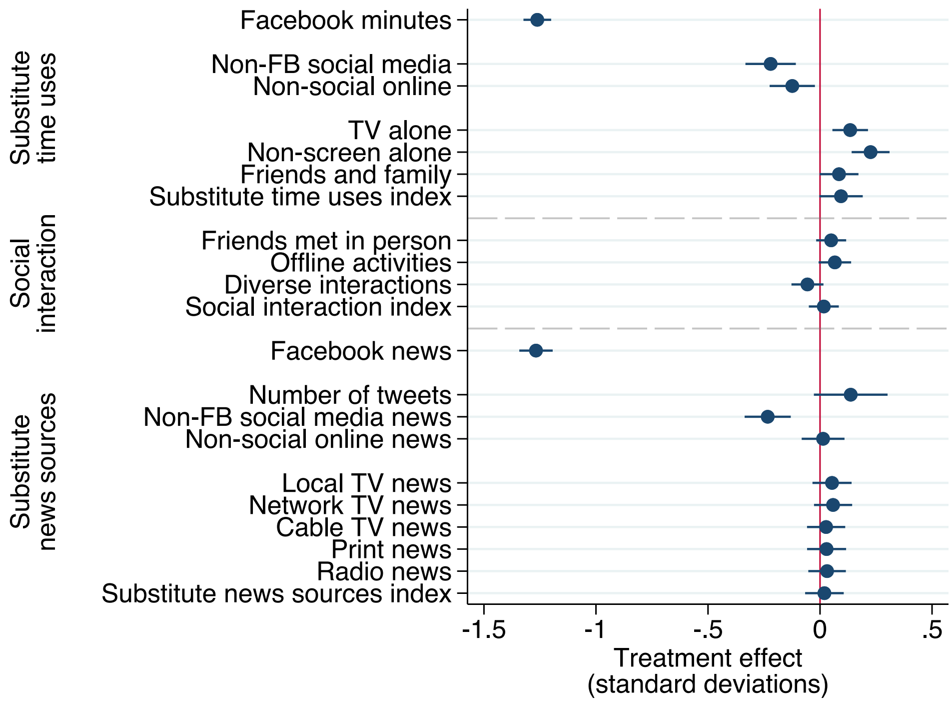
Table A9: **Treatment Effects: Post-Experiment Facebook Mobile App Usage**

| | (1) Full sample LATE | (2) Full sample ITT | (3) iPhone only LATE | (4) iPhone only ITT |
|---|---|---|---|---|
| Share of time deactivated | -12.15 | | -3.89 | |
| | (2.19) | | (2.78) | |
| Treatment | | -10.80 | | -3.52 |
| | | (2.00) | | (2.65) |
| Observations | 1,219 | 1,219 | 526 | 526 |
| Control group endline mean | 52.8 | 52.8 | 42.3 | 42.3 |
| Lee (2009) treatment effect lower bound | | -8.73 | | -2.04 |
| Lee (2009) treatment effect upper bound | | -7.76 | | -1.63 |
| Lee (2009) 95% confidence interval lower bound | | -13.77 | | -10.31 |
| Lee (2009) 95% confidence interval upper bound | | -3.18 | | 5.16 |

Notes: This table presents treatment effects of Facebook deactivation on post-experiment Facebook mobile app use in units of minutes per day, as measured in the December 3rd post-endline survey. Columns 1 and 2 include all observations, while columns 3 and 4 limit the sample to iPhone users who reported their Facebook mobile app usage as recorded by their System app, excluding participants who had reported personal estimates. Columns 1 and 3 present local average treatment effects estimated using Equation (1), while columns 2 and 4 present intent-to-treat effects and Lee (2009) bounds that account for attrition.

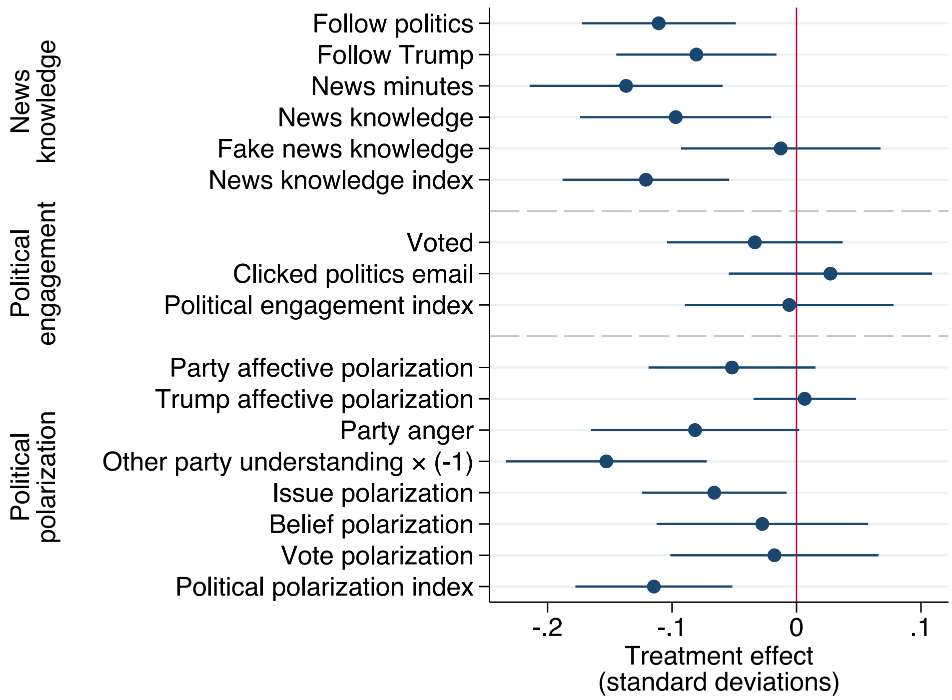
# D Treatment Effect Estimates Using Equation (2)

Figure A5: **Substitutes for Facebook Using Equation (2)**



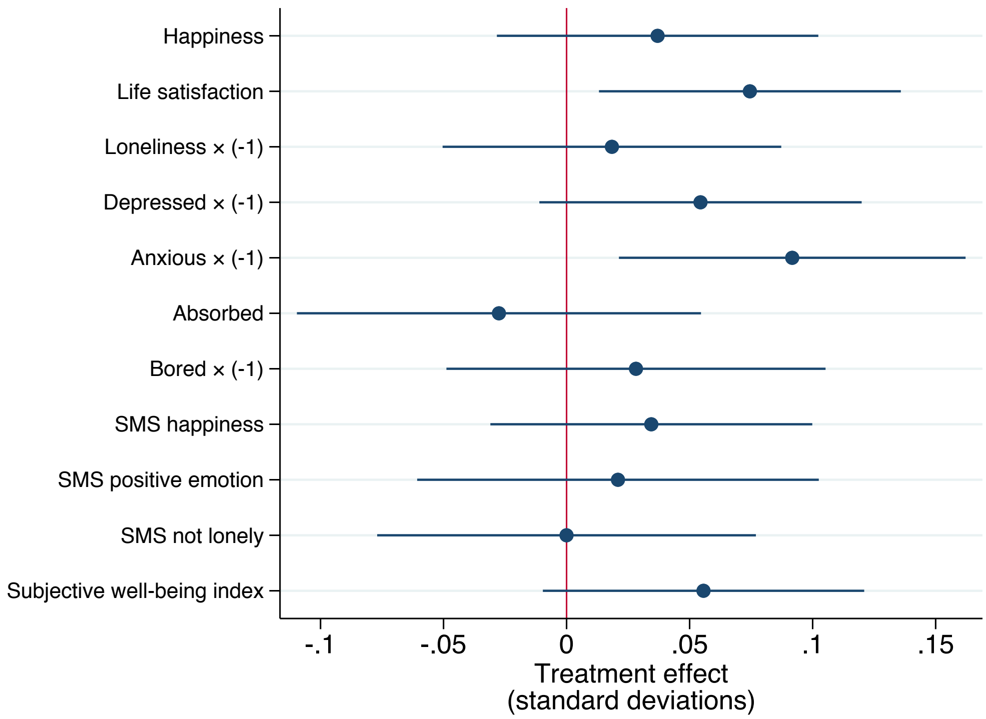

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (2). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A6: **Effects on News and Political Outcomes Using Equation (2)**



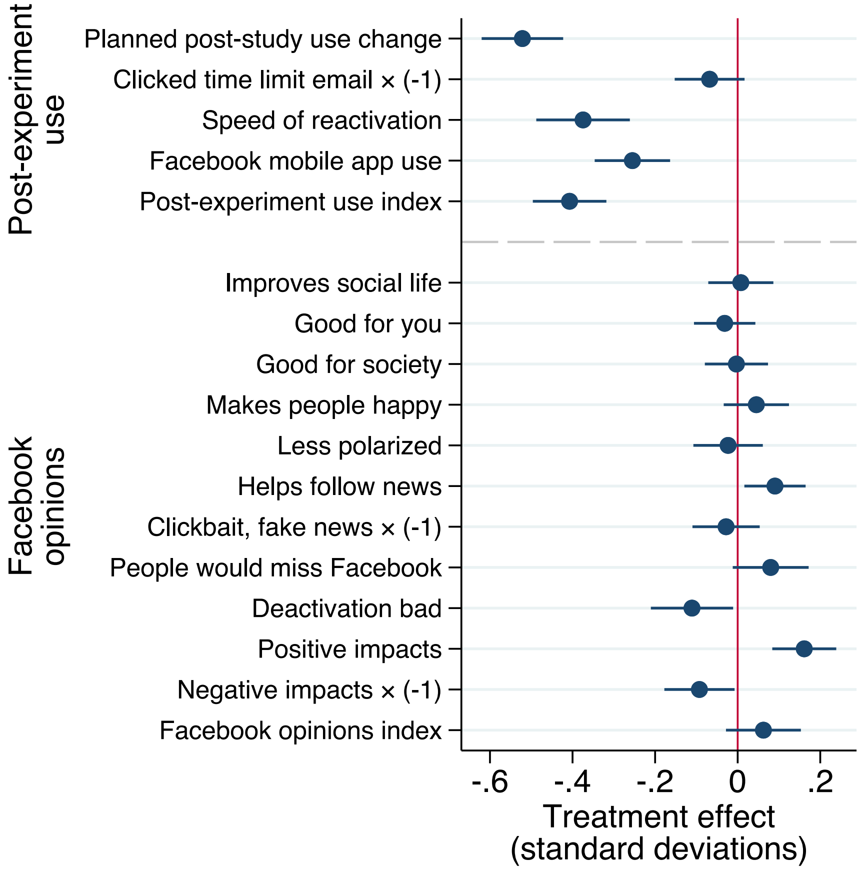

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (2). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A7: **Effects on Subjective Well-Being Using Equation (2)**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (2). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A8: **Effects on Post-Experiment Facebook Use and Opinions Using Equation** (2)

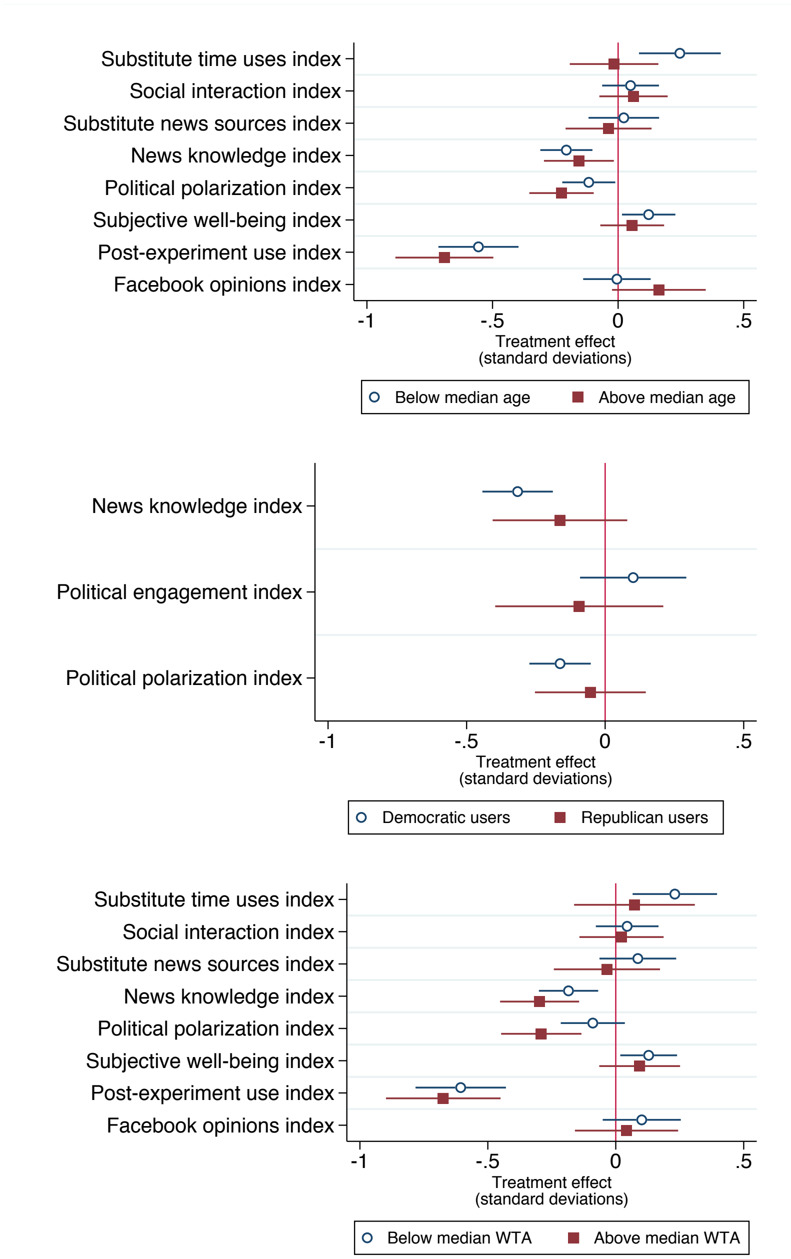



Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (2). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

# E Heterogeneous Treatment Effects

## E.1 Secondary Moderators

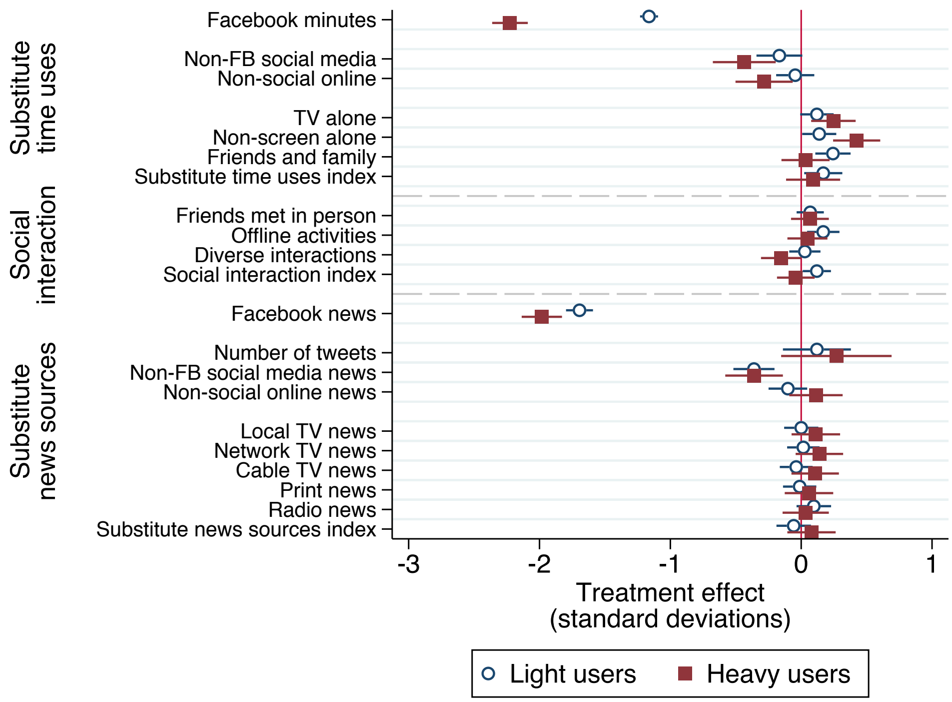
Figure A9: **Heterogeneous Treatment Effects for Secondary and Ex-Post Moderators**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). Age and political party were the "secondary" moderators in our pre-analysis plan. Willingness-to-accept was not defined as a moderator of interesting in our pre-analysis plan. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

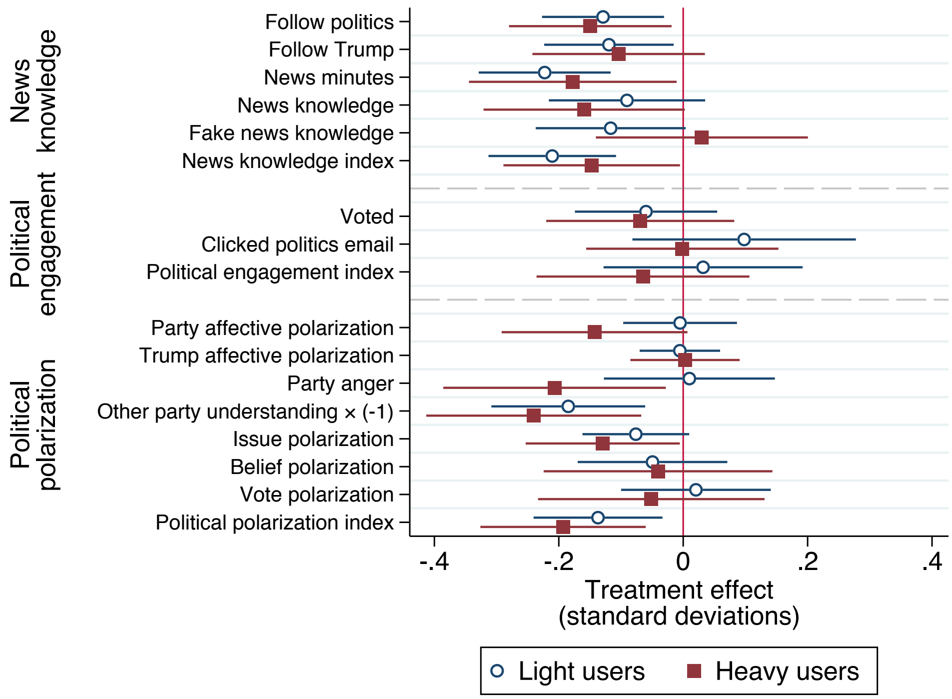
## E.2 Light and Heavy Users

Figure A10: **Substitutes for Facebook for Light and Heavy Users**



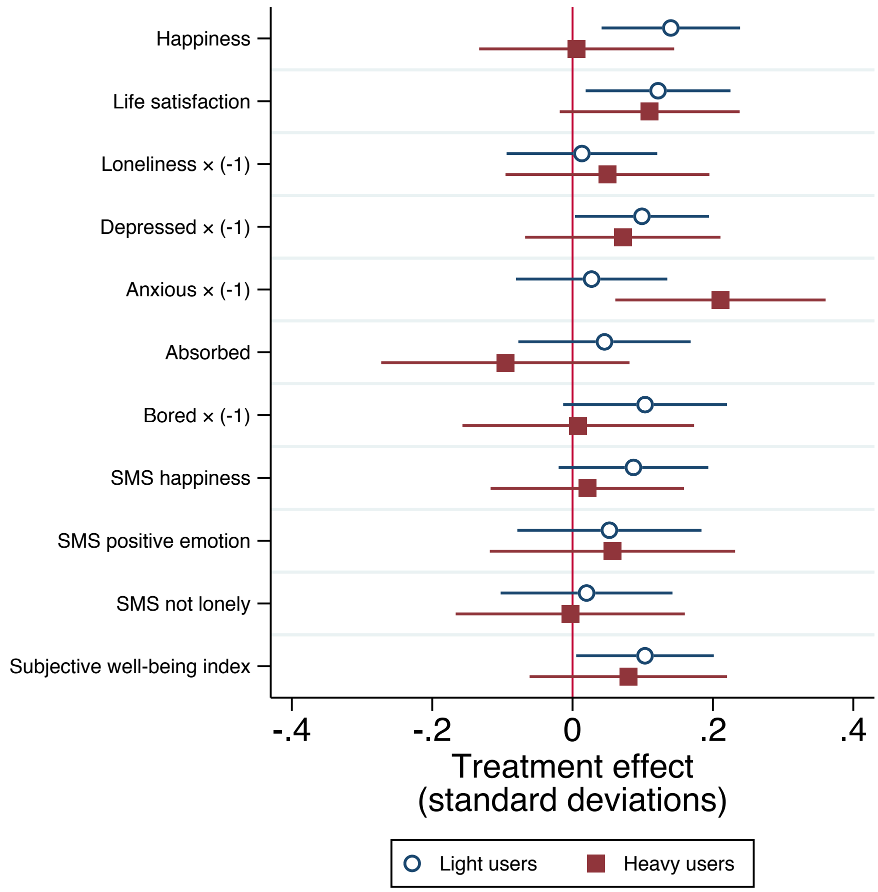

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 75 daily minutes, the median amount of Facebook use in the impact evaluation sample. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A11: **Effects on News and Political Outcomes for Light and Heavy Users**



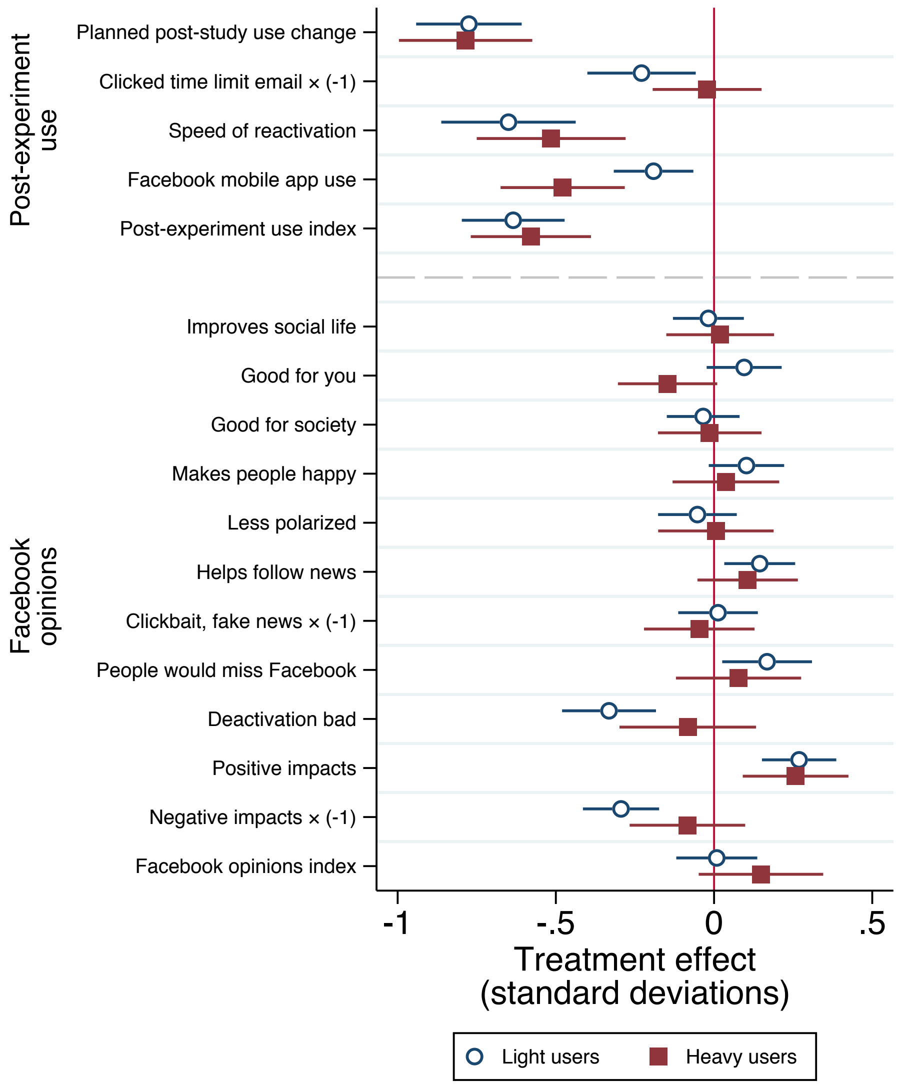

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 75 daily minutes, the median amount of Facebook use in the impact evaluation sample. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A12: **Effects on Subjective Well-Being for Light and Heavy Users**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 75 daily minutes, the median amount of Facebook use in the impact evaluation sample. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

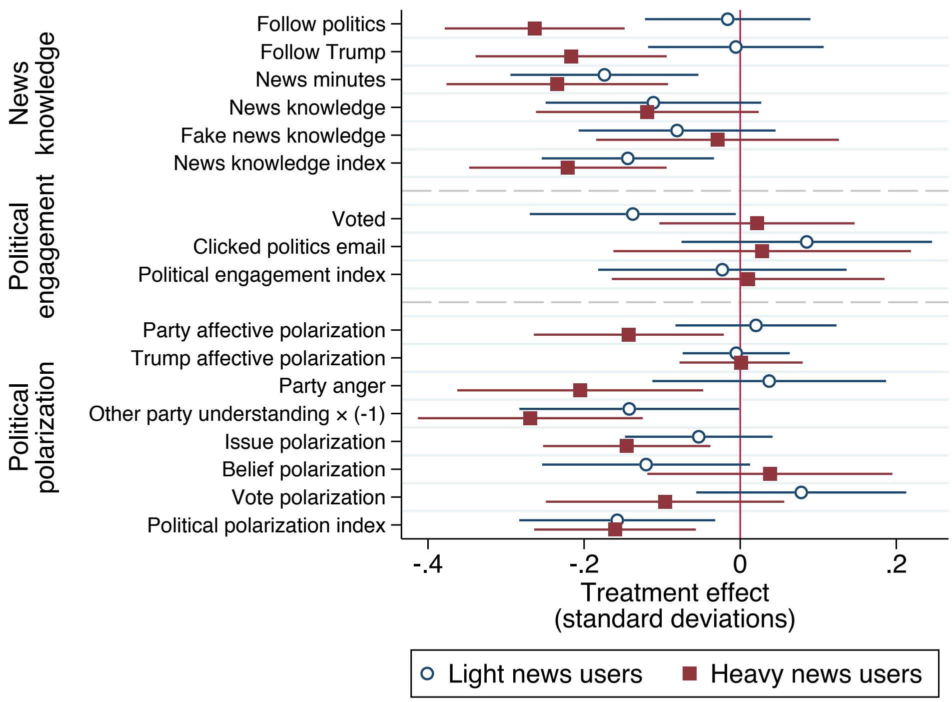
Figure A13: **Effects on Post-Experiment Facebook Use and Opinions for Light and Heavy Users**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 75 daily minutes, the median amount of Facebook use in the impact evaluation sample. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

## E.3 Light and Heavy News Users

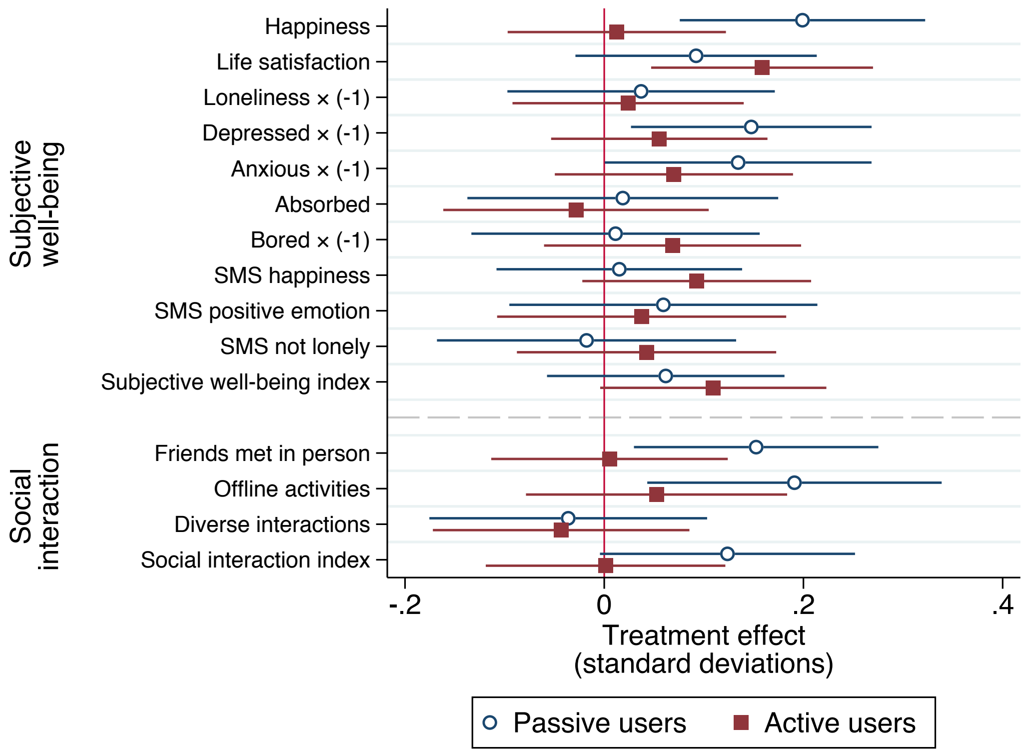
Figure A14: **Effects on News and Political Outcomes for Light and Heavy News Users**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for heavy news users vs. light news users (those who get news from Facebook fairly often or very often vs. never, hardly ever, or sometimes). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

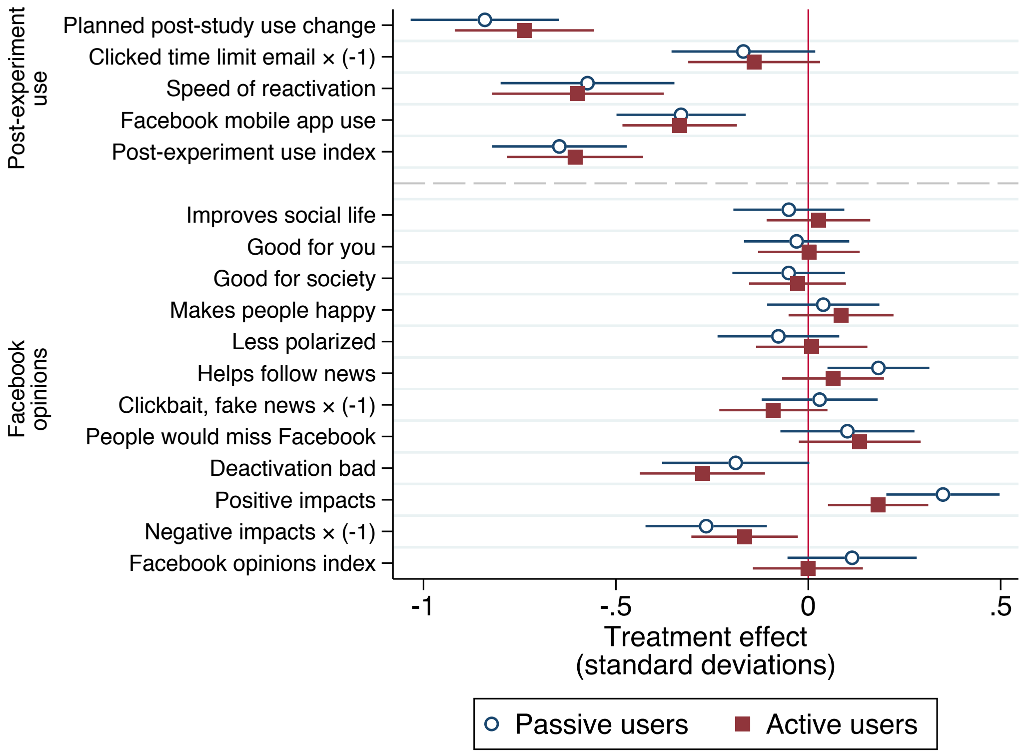
## E.4 Active and Passive Users

Figure A15: **Effects on Subjective Well-Being and Social Interactions for Active and Passive Users**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for active users vs. passive users. We measure this using two questions: share of active vs. passive browsing using a question based on the Passive and Active Facebook Use Measure (Gerson, Plagnol, and Corr 2017), and "what share of your time on Facebook do you spend interacting one-on-one with people you care about." Active vs. passive users are defined as having above- vs. below-median sum of their two responses to these questions. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

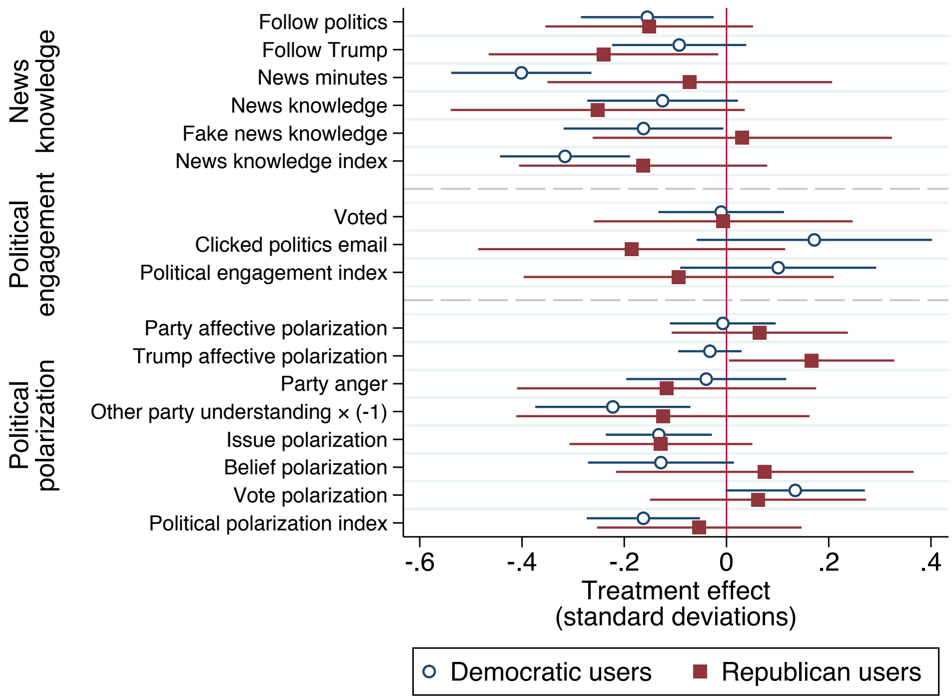
Figure A16: **Effects on Post-Experiment Use and Opinions about Facebook for Active and Passive Users**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for active users vs. passive users. We measure this using two questions: share of active vs. passive browsing using a question based on the Passive and Active Facebook Use Measure (Gerson, Plagnol, and Corr 2017), and "what share of your time on Facebook do you spend interacting one-on-one with people you care about." Active vs. passive users are defined as having above- vs. below-median sum of their two responses to these questions. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

## E.5 Democrats and Republicans

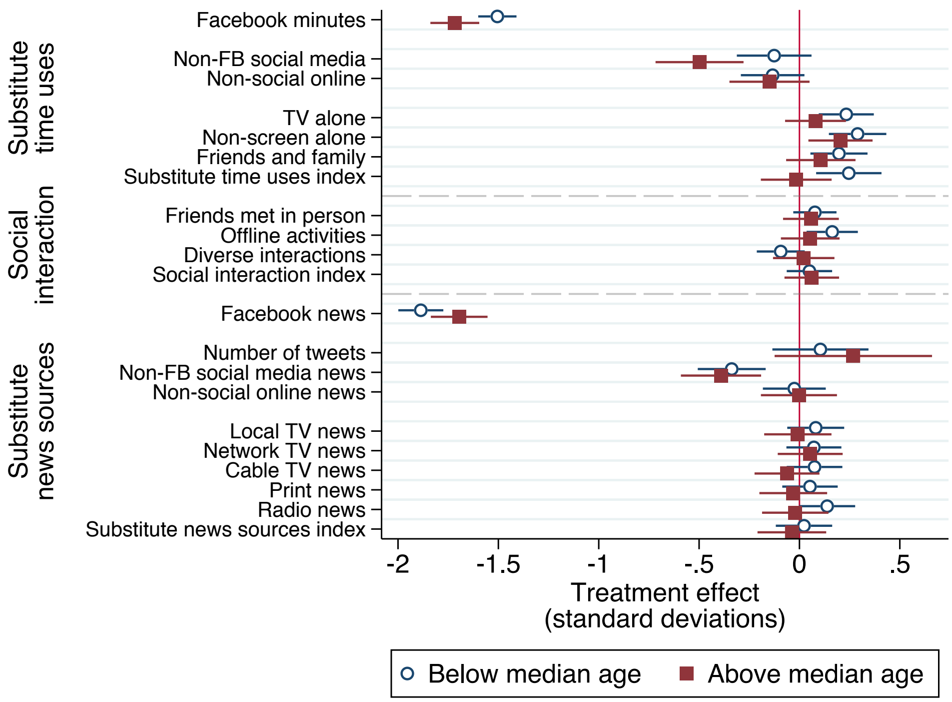
Figure A17: **Effects on News and Political Outcomes for Democrats and Republicans**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for Democrats vs. Republicans. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

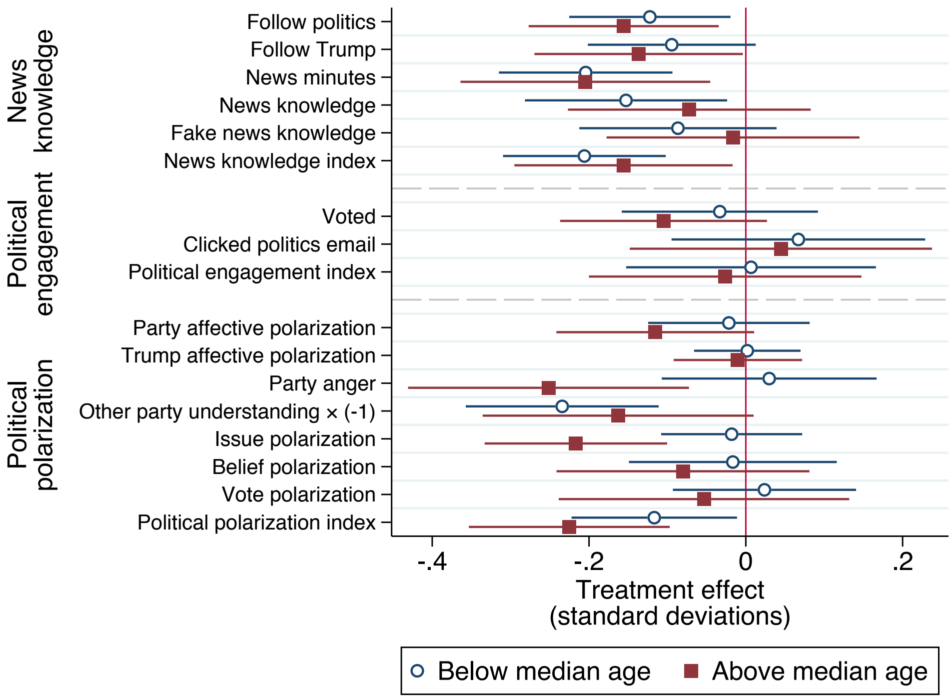
## E.6 Younger and Older Users

Figure A18: **Substitutes for Facebook for Younger and Older Users**



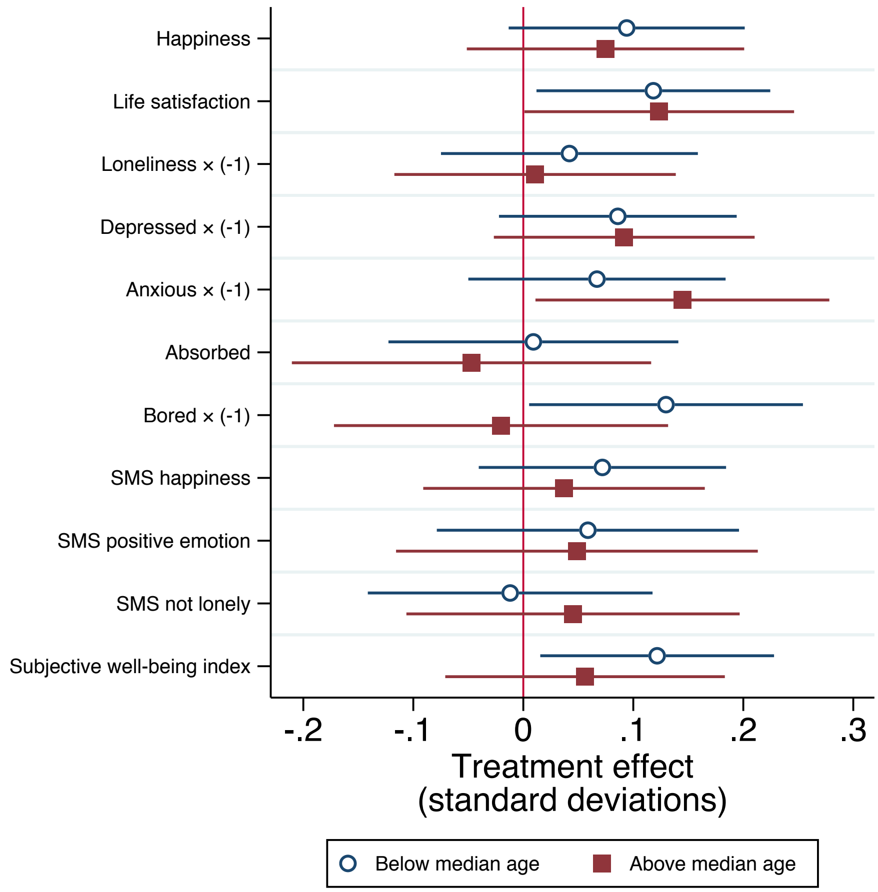

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 31.5 years, the median age in the impact evaluation sample. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A19: **Effects on News and Political Outcomes for Younger and Older Users**



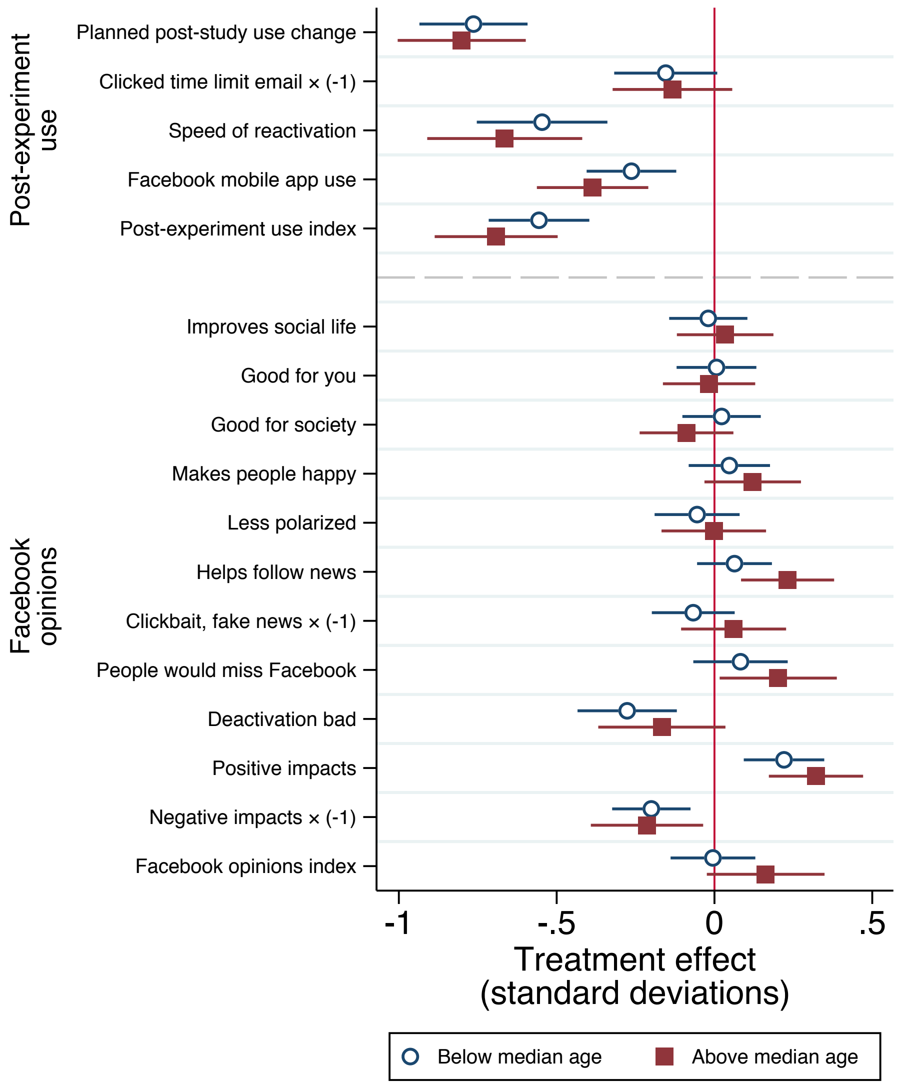

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 31.5 years, the median age in the impact evaluation sample. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A20: **Effects on Subjective Well-Being for Younger and Older Users**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 31.5 years, the median age in the impact evaluation sample. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A21: **Effects on Post-Experiment Facebook Use and Opinions for Younger and Older Users**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 31.5 years, the median age in the impact evaluation sample. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

# F News Knowledge

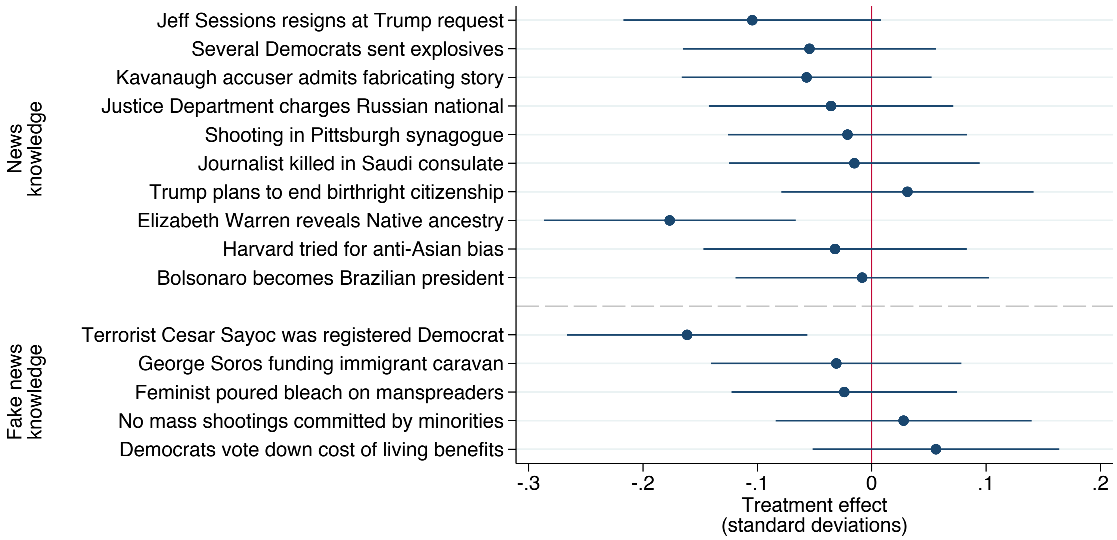
Appendix Figure A22 presents treatment effects on the probability of correct answers for each individual news knowledge question. Recall that we code a value of 1 for true statements correctly rated as true or incorrect statements correctly rated as false, 0.5 for any statement rated as "unsure," and 0 for true statements incorrectly rated as false or incorrect statements incorrectly rated as true.

To unpack these results, Appendix Figures A23, A24, and A25 present local average treatment effects of Facebook deactivation on indicators for answering true, false or unsure to our sets of true news, false news, and fake news questions respectively. By true news, we refer to the seven statements about news events reported by major outlets in which we did not insert factual inaccuracies. By false news, we refer to the three statements about news events reported by major news outlets in which we did insert substantial factual inaccuracies. By fake news, we refer to the five statements summarizing news articles that were deemed false on fact-checking websites and that circulated heavily within the four-week period before the survey. At the bottom of each block of news questions, we present treatment effects on the average across the questions in that block.

Most of the estimates are not statistically significant at any conventional level. Notwithstanding, the pattern of point estimates for true and false news statements is cohesive: in eight out of ten questions, deactivation induced people to move away from the correct answer and towards either the incorrect answer or "unsure" (or both). This paints a richer picture of how Facebook deactivation might reduce news knowledge: Treatment group participants are more likely to answer "unsure" and, if they do not answer "unsure" and take a guess as to whether the news event is true or false, they are more likely to answer incorrectly.

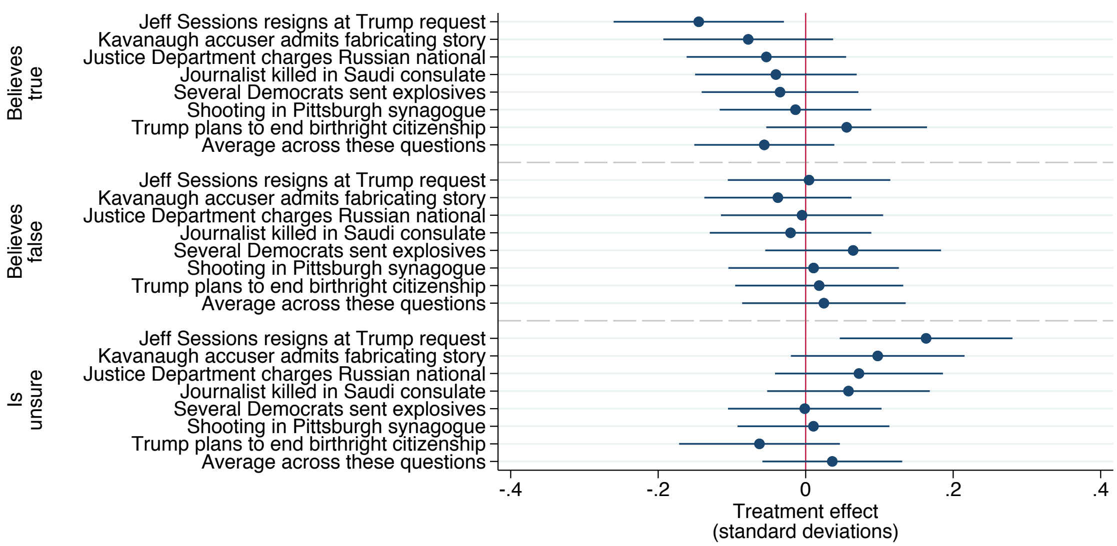
For the fake news quesetions, Facebook deactivation appears to have made people more likely to answer "unsure" instead of "false." This explains the negative point estimate of the effect of deactivation on fake news knowledge presented in Figure 3. Although not nearly statistically significant, one explanation for these point estimates is that Facebook circulates fake news but, at least for the major fake news stores in our survey, provides corrective information that helps users to correctly identify these stories as fake.

Figure A22: **Effects on News Knowledge and Fake News Knowledge**



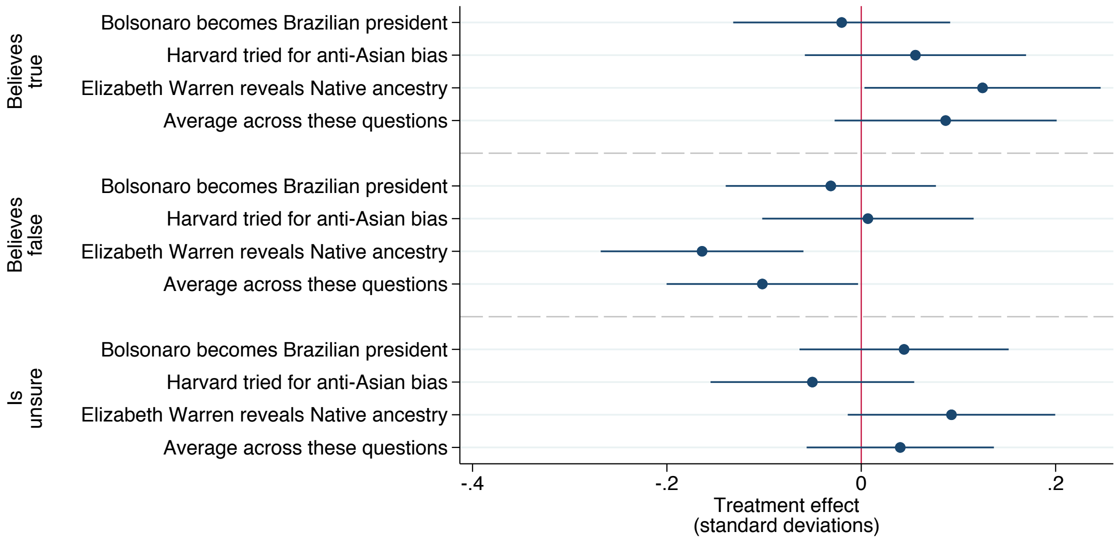

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A23: **Effects on Knowledge of True News Items**



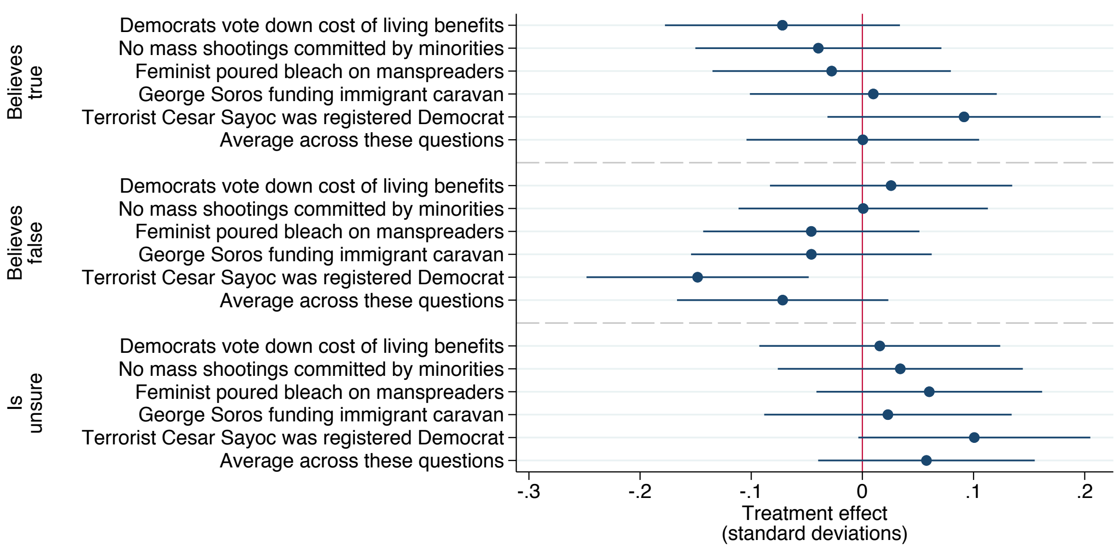

Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). The left-hand side variables are indicators for answering true, false or unsure to each of our true news items. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals.

Figure A24: **Effects on Knowledge of False News Items**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). The left-hand side variables are indicators for answering true, false or unsure to each of our false news items. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals.

Figure A25: **Effects on Knowledge of Fake News Items**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). The left-hand side variables are indicators for answering true, false or unsure to each of our fake news items. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals.

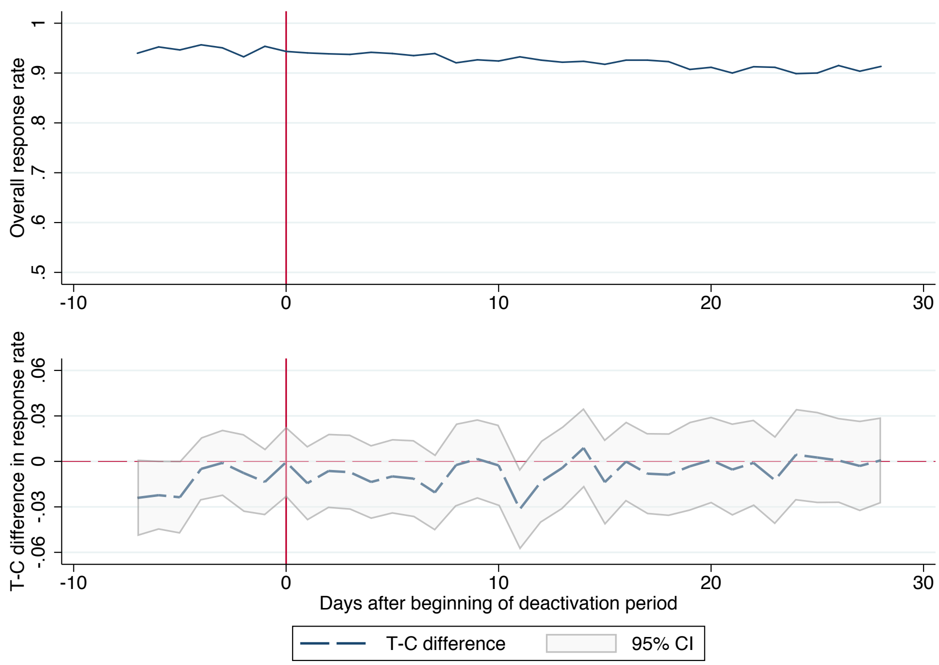
# G Additional Empirical Results

Table A10: **Balance**

| Variable | (1) Treatment Mean/SD | (2) Control Mean/SD | T-test P-value (1)-(2) |
|---|---|---|---|
| Income ($000s) | 71.27 | 72.69 | 0.59 |
| | (50.22) | (51.80) | |
| College | 0.52 | 0.50 | 0.61 |
| | (0.50) | (0.50) | |
| Male | 0.44 | 0.42 | 0.60 |
| | (0.50) | (0.49) | |
| White | 0.68 | 0.68 | 0.77 |
| | (0.47) | (0.46) | |
| Age | 33.04 | 32.34 | 0.27 |
| | (12.54) | (11.71) | |
| Republican | 0.13 | 0.14 | 0.85 |
| | (0.34) | (0.34) | |
| Democrat | 0.41 | 0.42 | 0.53 |
| | (0.49) | (0.49) | |
| Facebook minutes | 75.20 | 74.15 | 0.57 |
| | (35.58) | (35.49) | |
| Get news from Facebook | 3.47 | 3.43 | 0.45 |
| | (1.12) | (1.06) | |
| Active browsing | 0.14 | 0.16 | 0.73 |
| | (0.98) | (0.97) | |
| N | 580 | 1081 | |
| F-test of joint significance (p-value) | | | 0.95 |
| F-test, number of observations | | | 1661 |

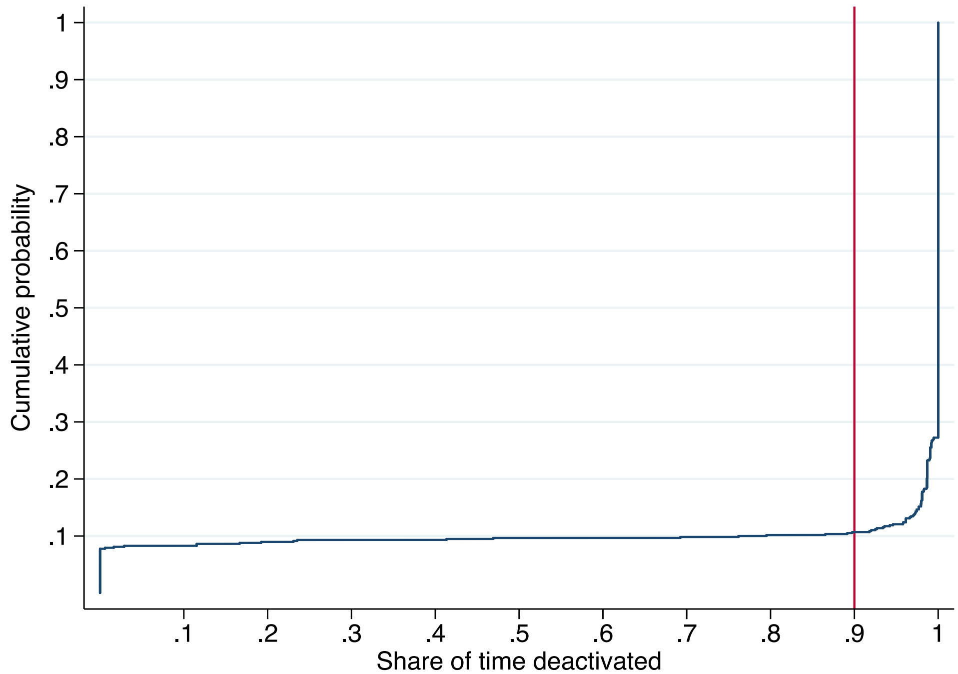
Notes: Columns 1 and 2 present demographics for the Treatment and Control groups in the impact evaluation sample: participants who were willing to accept less than $102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. Column 3 presents p-values of tests of differences in means between the two groups.

Figure A26: **Response Rates to Daily Text Messages**



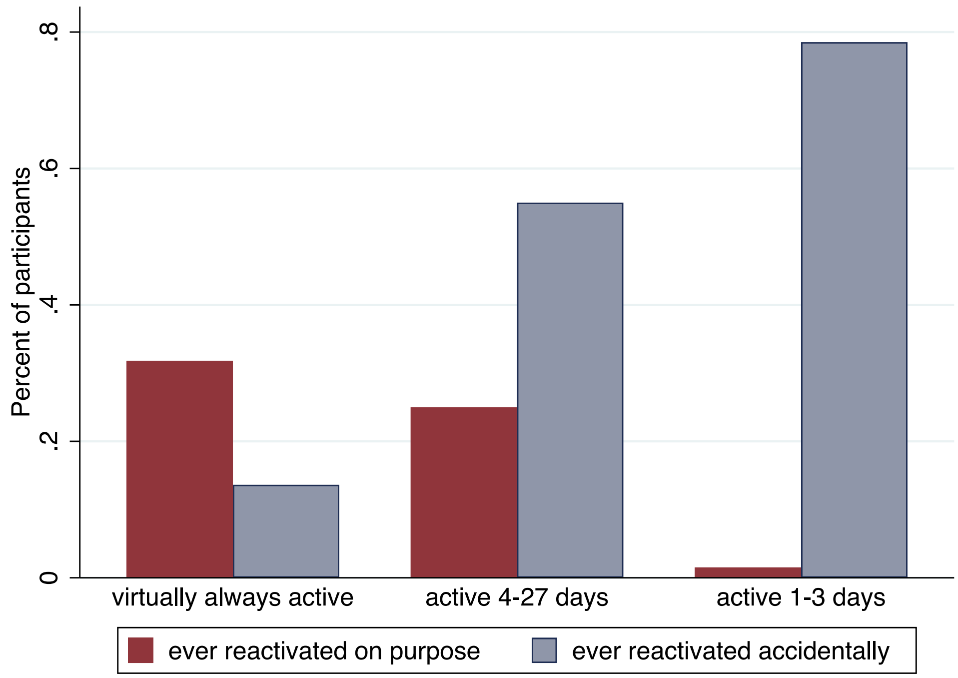

Notes: The figure shows response rates to the SMS survey and the difference in response rates between Treatment and Control, for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered $p = \$102$ or $p = \$0$ to do so. The vertical red line reflects the date of the midline survey.

Figure A27: **Treatment Group Distribution of Share of Time Deactivated**



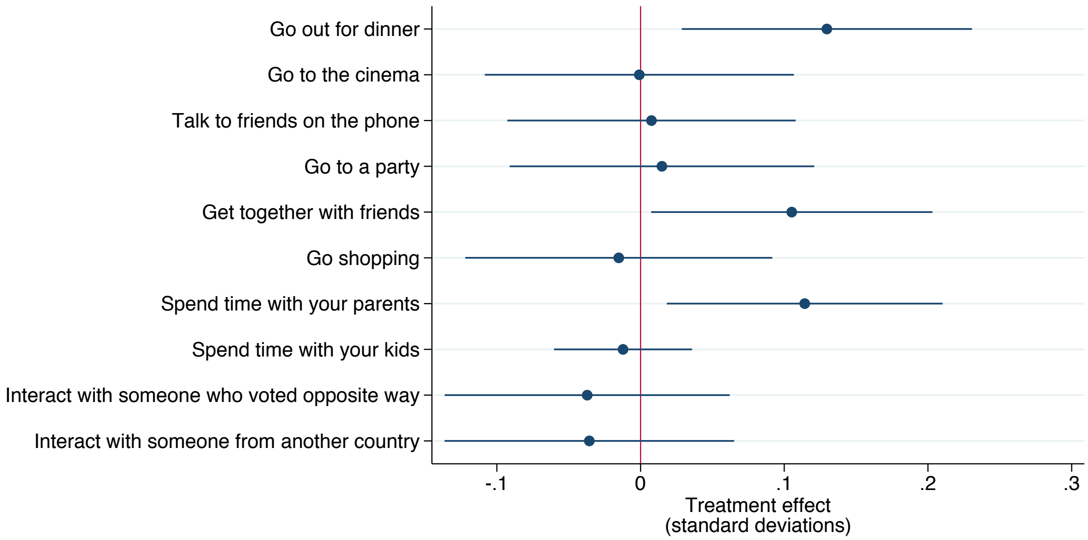

Notes: For each individual in the Treatment group who was willing to accept less than $102 to deactivate Facebook for the four weeks after midline, we calculate the share of the deactivation checks in which that person was deactivated. This figure presents the cumulative distribution of the share of the time deactivated across people.

Figure A28: **Reasons for Failure to Deactivate**



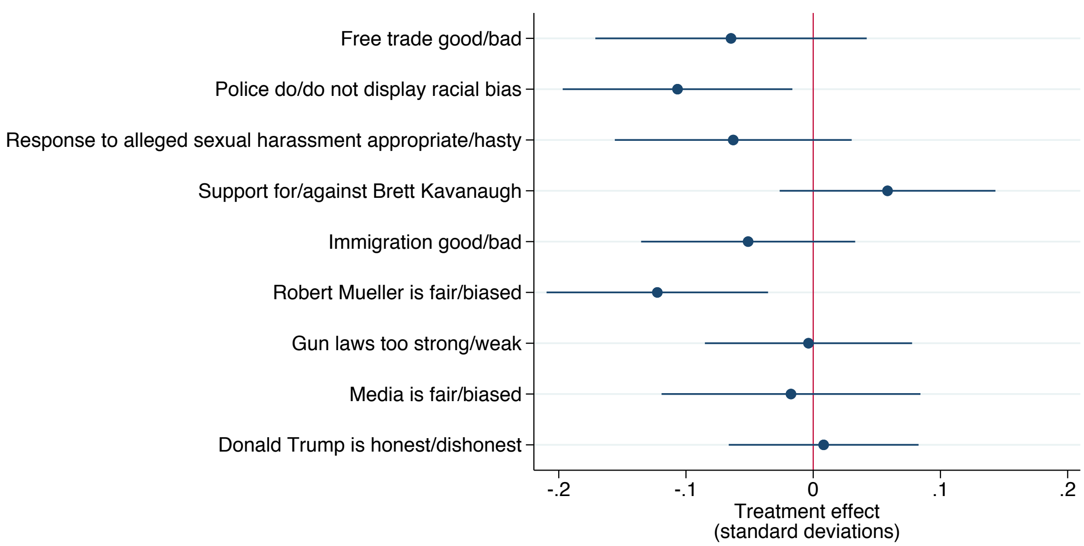

Notes: This figure presents reasons for failure to deactivate for Treatment group participants. Data were gathered from an optional survey that we emailed to participants who were not deactivated when they were supposed to be under the experiment protocols. The survey asked, "Why did your Facebook account get reactivated? Your answer won't affect your payment – we're just trying to figure out what problems people are having." Possible responses were, "I logged into my account using the Facebook website or the Facebook app," "somebody else logged into my account," "I used an app (other than the Facebook app or the Facebook messenger app) that uses my Facebook credentials to log in," "Other (please specify)," and "I don't know." We coded an individual as having reactivated "on purpose" if they ever clicked the first answer ("I logged into my account"). We coded an individual as having reactivated "accidentally" if they ever clicked on the second, third, or fifth answers. We also manually coded text that respondents wrote in the "Other (please specify)" box as either "on purpose" or "accidental."

Figure A29: **Effects on Offline Activities and Diverse Interactions**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A30: **Effects on Issue Polarization**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Table A11: **Robustness to Omitting Each Individual Variable from the Political Polarization Index**

| | Treatment effect | Standard error | P-value |
|---|---|---|---|
| Party affective polarization | -0.15 | 0.04 | 0.00 |
| Trump affective polarization | -0.15 | 0.04 | 0.00 |
| Party anger | -0.15 | 0.04 | 0.00 |
| Other party understanding × (-1) | -0.07 | 0.04 | 0.06 |
| Issue polarization | -0.14 | 0.05 | 0.00 |
| Belief polarization | -0.11 | 0.04 | 0.00 |
| Vote polarization | -0.16 | 0.04 | 0.00 |

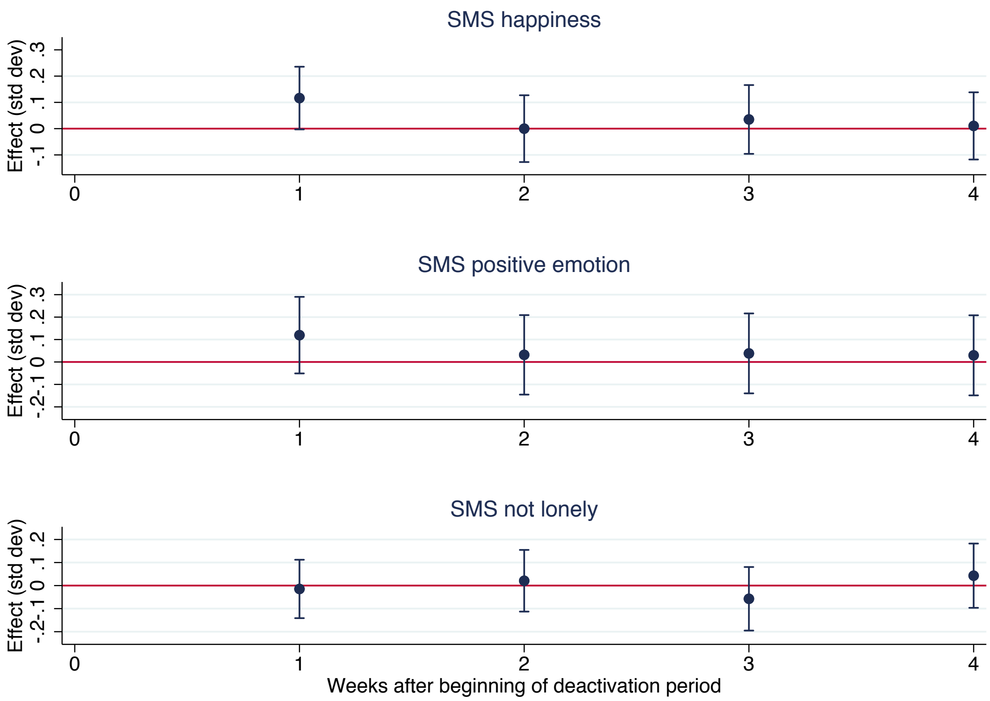
Notes: This table presents local average treatment effects of Facebook deactivation on the political polarization index estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Each row omits the variable listed from the index. See Section 2.3 for variable definitions.

Table A12: **Correlation Between Subjective Well-Being Index and Demographics at Baseline**

| | (1) |
|---|---|
| Income ($000s) | 0.0027 |
| | (0.0005) |
| College | 0.2335 |
| | (0.0488) |
| Male | 0.2033 |
| | (0.0482) |
| White | -0.0066 |
| | (0.0531) |
| Age | 0.0154 |
| | (0.0021) |
| Republican | 0.2136 |
| | (0.0723) |
| Democrat | -0.0492 |
| | (0.0507) |
| Observations | 1,661 |

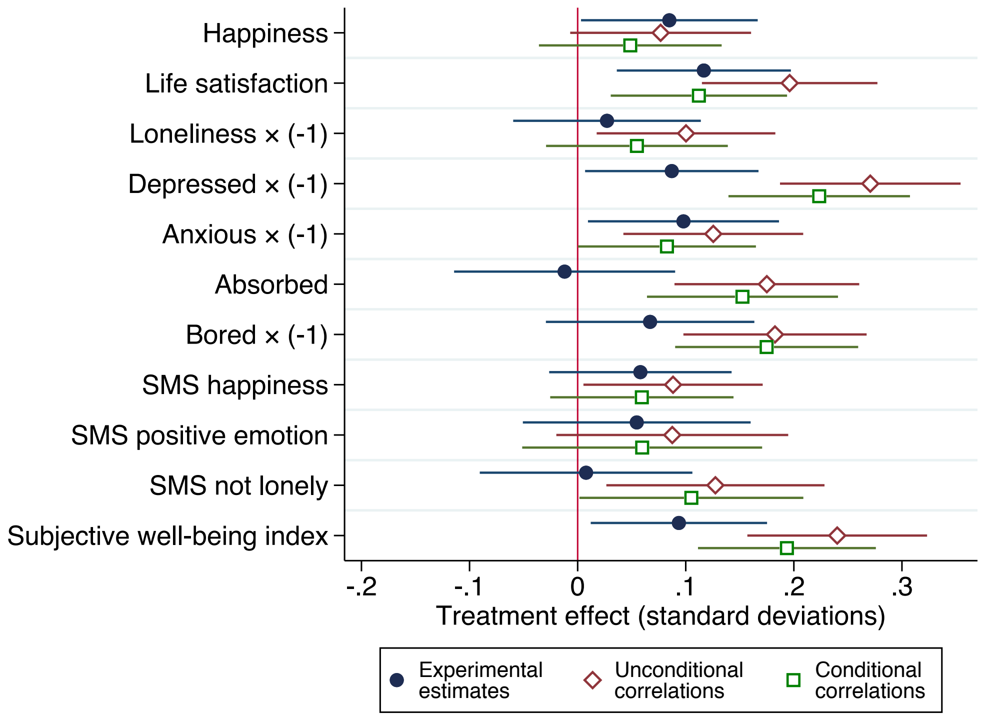
Notes: This table presents estimates of a regression of the baseline subjective well-being index on demographic variables. The subjective well-being index is normalized to have a standard deviation of one.

Figure A31: **Effects on Subjective Well-Being Measured in Text Messages, By Week**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A32: **Comparing Experimental and Non-Experimental Estimates of Effects on Subjective Well-Being**

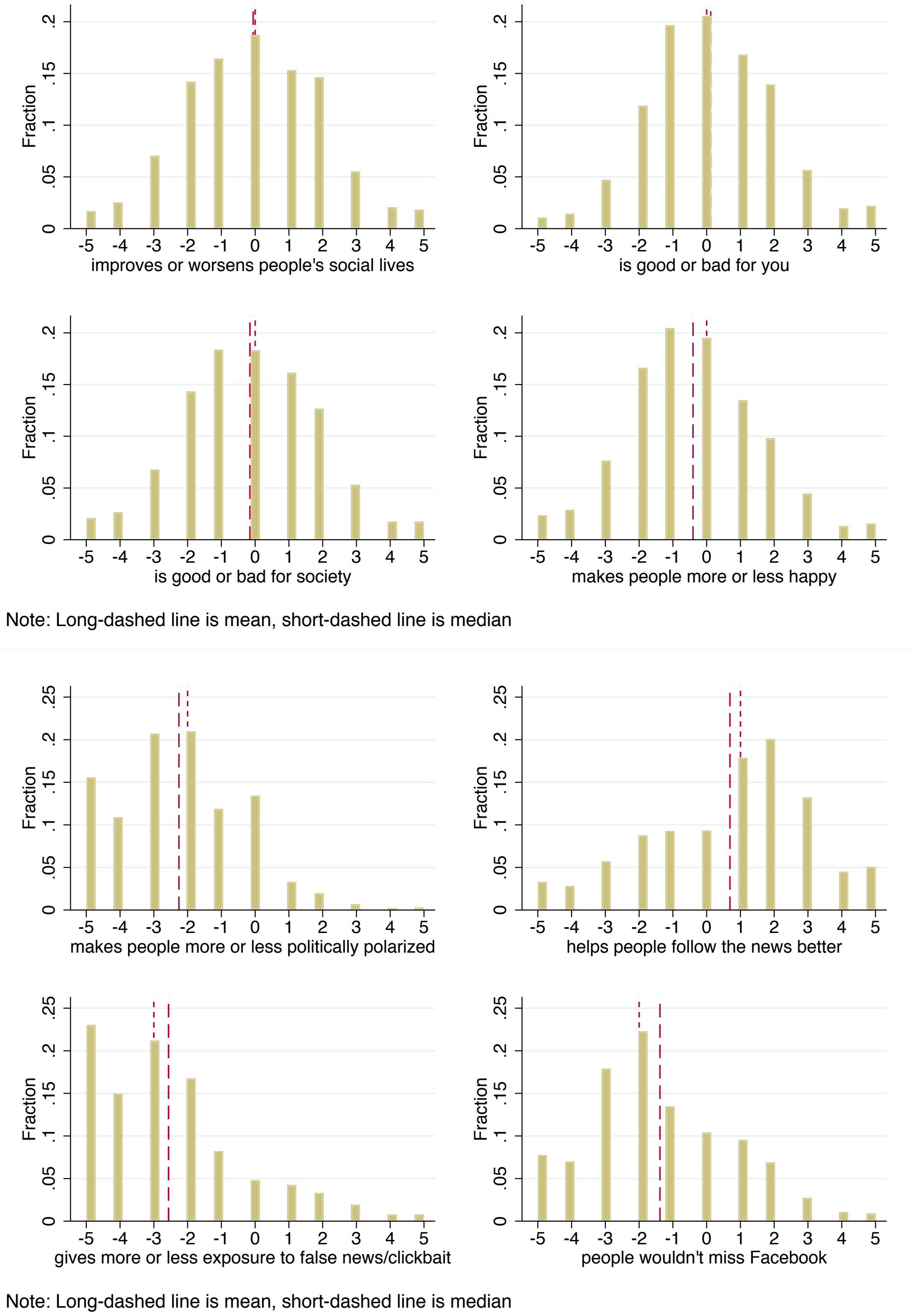



Notes: The solid markers present local average treatment effects of Facebook deactivation estimated using Equation (1). The empty markers present non-experimental estimates from the following regression:

$$Y_i^b = \tau \tilde{H}_i + \beta \boldsymbol{X}_i + \epsilon_i,$$

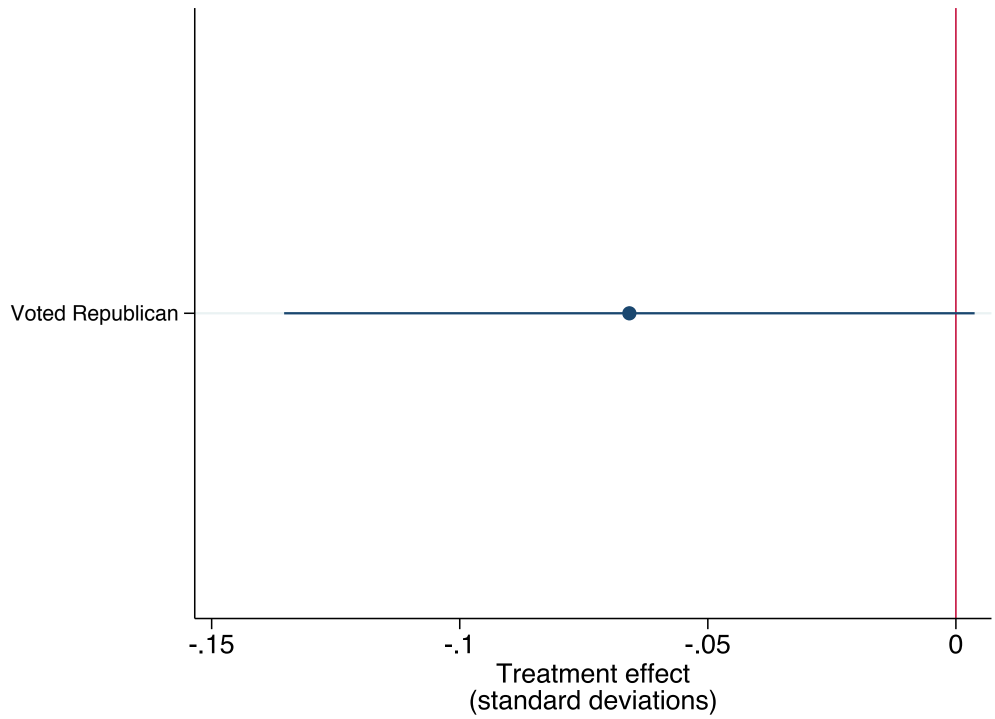
where $Y_i^b$ is participant $i$'s value of some outcome measured in the baseline survey, $\boldsymbol{X}_i$ is a vector of controls (household income, age, and college, male, white, Republican, and Democrat indicators), and $\tilde{H}_i$ is baseline average daily Facebook use over the past four weeks (winsorized at 120 minutes per day) divided by the local average treatment effect on average daily Facebook use between midline and endline. This division makes experimental and non-experimental estimates comparable in the sense that they are both in units of average use per day over the past four weeks. The empty diamond markers present unconditional correlations (excluding $\boldsymbol{X}_i$ from the regressions), while the empty square markers present estimates conditional on $\boldsymbol{X}_i$. All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A33: **Baseline Opinions about Facebook**



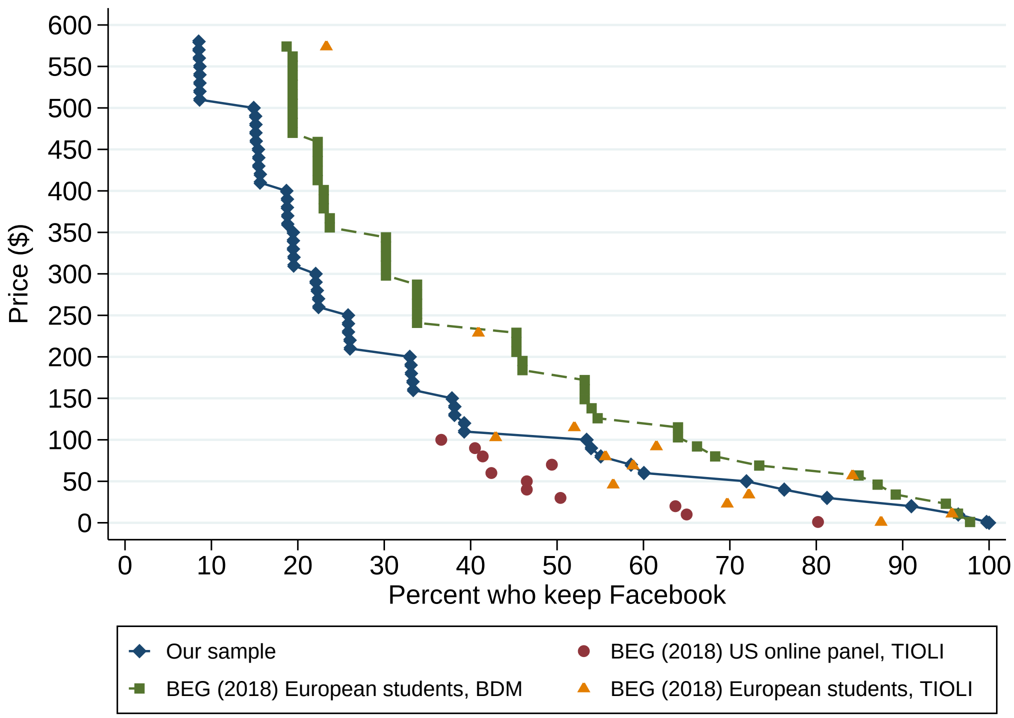

Notes: These figures present histograms of Facebook opinions from the baseline survey. Variables are re-signed so that "positive" views about Facebook are positive, "negative" views about Facebook are negative, and zero is neutral. See Section 2.3 for variable definitions.

Figure A34: **Effects on Secondary Outcomes**




Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the Control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section 2.3 for variable definitions.

Figure A35: **Comparison to Demand Curves from Brynjolffson et al. (2018)**




Notes: This figure compares our demand curve (based on the distribution of willingness-to-accept to deactivate for the four weeks after midline) to demand curves for one month of Facebook use from Brynjolfsson, Eggers, and Gannamaneni (2018). "TIOLI" refers to their "take it or leave it" elicitation, whereas "BDM" refers to their BDM elicitation. For their European student sample, valuations were elicited in Euros; we transform these to dollars using the exchange rate when the elicitation was carried out in July 2017.

# H Model Appendix

## H.1 Derivations of Facebook Valuations

For these derivations, denote $\boldsymbol{c} = \{c_1, c_2, ...., c_T\}$ as the composite good consumption vector, $\boldsymbol{f} = \{f_{1,}, f_2, ..., f_T\}$ as the Facebook consumption vector, and write $\tilde{U}_1(\cdot; f_{\tau-1})$ more precisely as $\tilde{U}_1(\boldsymbol{c}, \boldsymbol{f}; f_{\tau-1})$.

### WTA for month 1 deactivation as of midline

Find the $v_{1,1}(1, 1)$ that equates perceived utility with Facebook to perceived utility without Facebook:

$$\tilde{U}_1(\boldsymbol{c}, \{1,1,1,...\}; 1) = \tilde{U}_1(\boldsymbol{c}, \{0,1,1,....\}; 1) + v_{1,1}(1,1)$$

$$W + \underbrace{\phi + \alpha + \xi + \omega_1}_{t=1} + \underbrace{\phi + \alpha + \xi}_{t=2} + \sum_{t=3}^{T} (\phi + \alpha + \xi) = W + \underbrace{0}_{t=1} + \underbrace{\alpha + \xi}_{t=2} + \sum_{t=3}^{T} (\phi + \alpha + \xi) + v_{1,1}(1)$$

$$v_{1,1}(1,1) = \phi + \alpha + \xi + \omega_1 + \phi$$

**Treatment group: WTA for month 2 deactivation as of midline**

$$\tilde{U}_1(\boldsymbol{c}, \{0,1,1,...\}; 1) = \tilde{U}_1(\boldsymbol{c}, \{0,0,1,...\}; 1) + v_{2,1}(0,1)$$

$$W + \underbrace{0}_{t=1} + \underbrace{\alpha + \xi}_{t=2} + \underbrace{\phi + \alpha + \xi}_{t=3} + \sum_{t=4}^{T} (\phi + \alpha + \xi) = W + \underbrace{0}_{t=1} + \underbrace{0}_{t=2} + \underbrace{\alpha + \xi}_{t=3} + \sum_{t=4}^{T} (\phi + \alpha + \xi) + v_{2,1}(0,1)$$

$$v_{2,1}(0,1) = \alpha + \xi + \phi$$

**Control: WTA for month 2 deactivation as of midline**

$$\tilde{U}_1(\boldsymbol{c}, \{1,1,1,...\}; 1) = \tilde{U}_1(\boldsymbol{c}, \{1,0,1,...\}; 1) + v_{2,1}(1,1)$$

$$W + \underbrace{\phi + \alpha + \xi + \omega_1}_{t=1} + \underbrace{\phi + \alpha + \xi}_{t=2} + \underbrace{\phi + \alpha + \xi}_{t=3} + \sum_{t=4}^{T} (\phi + \alpha + \xi) = W + \underbrace{\phi + \alpha + \xi + \omega_1}_{t=1} + ...$$

$$... \underbrace{0}_{t=2} + \underbrace{\alpha + \xi}_{t=3} + \sum_{t=4}^{T} (\phi + \alpha + \xi) + v_{2,1}(1,1)$$

$$v_{2,1}(1,1) = \phi + \alpha + \xi + \phi$$

**Treatment: WTA for month 2 deactivation as of endline**

$$\tilde{U}_2(\boldsymbol{c}, \{0,1,1,...\}; 0) = \tilde{U}_2(\boldsymbol{c}, \{0,0,1,....\}; 0) + v_{2,2}(0,0)$$

$$W + \underbrace{\xi + \omega_2}_{t=2} + \underbrace{\phi + \xi}_{t=3} + \sum_{t=4}^{T} (\phi + \xi) = W + \underbrace{0}_{t=2} + \underbrace{\xi}_{t=3} + \sum_{t=4}^{T} (\phi + \xi) + v_{2,2}(0,0)$$

$$v_{2,2}(0,0) = \xi + \omega_2 + \phi$$

**Control: WTA for month 2 deactivation as of endline**

$$\tilde{U}_2(\boldsymbol{c}, \{1,1,1,...\};1) = \tilde{U}_2(\boldsymbol{c}, \{1,0,1,....\};1) + v_{2,2}(1,1)$$

$$W + \underbrace{\phi + \alpha + \xi + \omega_2}_{t=2} + \underbrace{\phi + \alpha + \xi}_{t=3} + \sum_{t=4}^{T} (\phi + \alpha + \xi) = W + \underbrace{0}_{t=2} + \underbrace{\alpha + \xi}_{t=3} + \sum_{t=4}^{T} (\phi + \alpha + \xi) + v_{2,2}(1,1)$$

$$v_{2,2}(1,1) = \phi + \alpha + \xi + \omega_2 + \phi$$