The Welfare Effects of Social Media

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

December 17, 2018

Abstract

The rise of social media has provoked both optimism about potential benefits to society, and concern about harms ranging from addiction to depression to political polarization. We present a large-scale, randomized evaluation of the welfare impacts of Facebook, focusing on US users in the run-up to the 2018 midterm election. We measure the willingness to accept of 2,844 Facebook users to deactivate their Facebook accounts for a period of four weeks, then randomly assign a subset to actually do so in a way that we verify. Using a suite of outcomes from both surveys and direct measurement, we show that Facebook deactivation (i) reduces online activity including other social media, while increasing offline activities including time with family and friends; (ii) has precisely estimated zero effects on subjective well being; (iii) reduces both factual knowledge of news and political polarization; and (iv) causes a persistent reduction in use of Facebook following the experiment. We combine these facts with the incentivized valuations we elicit to calibrate a simple behavioral model. We conclude that forces like addiction and mis-prediction may lead to some over-consumption, but that the magnitude of these mistakes are likely small relative to the total consumer surplus gains from Facebook.

---

*Allcott: New York University, NBER, and JPAL. hunt.allcott@nyu.edu. Braghieri: Stanford University. lucabrag@stanford.edu. Eichmeyer: Stanford University. saraeich@stanford.edu. Gentzkow: Stanford University and NBER. gentzkow@stanford.edu. We thank Nancy Baym, Moira Burke, Alex Leavitt, and Todd Rogers for helpful comments. We thank Kelly Wagman and Raj Bhargava for exceptional research assistance. We are grateful to the Sloan Foundation for grant funding. The study was approved by IRBs 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. Replication files and 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.

I Introduction

The rise of social media has had a profound impact around the world. The current number of monthly Facebook users stands at 2.3 billion worldwide, and as of 2016 the average user was spending 50 minutes per day on Facebook plus its sister platforms Instagram and Messenger. 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, 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.

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 (Twenge, Sherman, and Lyubomirsky 2016; Twenge and Park 2017; Twenge, Martin, and Campbell 2018; Twenge et al. 2018). Alter (2018), 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). 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 3,051 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 55 percent of these subjects with valid WTA measures less than $102 to either a Treatment group which was paid to deactivate, or a Control group which was not. We verified compliance with deactivation using public profile pages. Over the course of the experiment, we

[1] See, for example, Burke and Kraut 2016, 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 Crusiu 2016 and Baker and Algorta 2016 for reviews.

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.

We offer the largest scale experimental evidence available to date on the way Facebook impacts a range of individual and social welfare measures. We evaluate the extent to which time on Facebook substitutes for alternative activities both online and offline, with particular attention to crowd out of news consumption and face to face social interactions. We study Facebook's impact on individual utility via measures of subjective well-being and social connectedness captured through both surveys and text messages. We study broader political externalities via measures of news knowledge, political engagement, awareness of misinformation, and political polarization. Finally, we analyze the extent to which behavioral forces like addiction and mis-prediction may lead individuals to make sub-optimal consumption choices, looking at how both WTA and actual Facebook usage change after our experiment, and comparing this to participants own *ex ante* predictions.

Attrition during the experiment was low, and what attrition there was was balanced between Treatment and Control. Our endline survey was completed by more than 98 percent of both the Treatment and Control group, and our text messages and post-endline surveys are completed by more than 90 percent of subjects. Levels of compliance by the Treatment group in our experiment were high. We randomly checked Treatment subjects' activation status approximately twice per day, and more than 90 percent of these checks showed the accounts were indeed inactive. In contrast, only 2 percent of the Control group accounts we checked were inactive. Follow-up analysis shows that the 10 percent activation in Treatment was a combination of persistent, intentional non-compliance by a small group of participants and brief inadvertent reactivations among the remainder, for example because they logged into another website using their Facebook credentials.

Our first set of results focuses on substitution patterns. A key mechanism for effects on individual well being is social media crowding out face-to-face social interactions and thus deepening loneliness and depression. A key mechanism for political externalities is social media crowding out consumption of higher quality news and information sources sources. We find evidence consistent with the first of these but not with the second. Facebook deactivation causes our Treatment group to reduce time on all online activities including non-Facebook social media and other forms of online media, and to increase time on a broad range of offline activities including time with friends and family. Their total time spent reading news falls, and we do not see significant increases in consumption of any online or offline news sources.

Our second set of results looks at individual utility as captured by subjective well-being and reported social interactions. Here, we find precise zero effects in every case, including standard measures of happiness, life satisfaction, loneliness, depression, anxiety, and other outcomes, measured both in the endline survey and in daily text messages. Using our overall index of subjective well-being, we can rule out negative effects of Facebook use of even 0.1 standard deviations with 95 percent confidence. These results argue against the view that Facebook is reducing individual

welfare to a large extent, and may suggest caution in interpreting prior correlational studies. We find no evidence for the hypothesis from prior research that "passive" Facebook use—for example, reading through your news feed instead of commenting on pictures and posts from your friends and family—is more strongly associated with negative psychological outcomes (e.g. Aalbers et al. 2018). We find no differential impacts of deactivation on subjective well-being for participants who reported being "passive" Facebook users before the experiment.

Our third set of results turns to broader social impacts. Consistent with the substitution patterns we observe, we find that deactivation of Facebook significantly reduces news knowledge and attention to politics. Our treatment group is less likely to say they follow politics or the President, and less able to answer factual questions about recent news events. Our overall index of news knowledge falls by 0.2 standard deviations. The point estimate on voting in the midterm election is insignificantly negative, and we find no effect on the likelihood of clicking email links to support political causes. At the same time, we find that deactivation also significantly reduces political polarization, particularly as measured by attitudes on policy issues and inter-party understanding or anger. Our overall index of polarization falls by 0.15 standard deviations.

Our final set of results provides a more holistic look at the extent to which behavioral forces may lead people to make sub-optimal consumption choices. A key piece of evidence in this section is that our treatment group reduces consumption of Facebook following the experiment by a number of measures. Treatment users' self-reported intention to use Facebook in the future is lower, their actual usage as measured in our post-endline survey is lower (though this falls just shy of significance), and a significant share keep their accounts deactivated for weeks after the experiment has ended. Treatment users are also more likely to click on a post-endline email providing information about tools to limit one's own social media usage. Overall, our index of post-experiment Facebook use falls by more than half a standard deviation.

Our results should be interpreted with caution, for several reasons. First, the effects of a longer deactivation could be larger or smaller, depending on the outcome. A longer period without Facebook might have less impact on news knowledge as people find alternative news sources, and more impact on subjective well-being. Second, we have standard concerns about sample selection bias. Our sample is 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. Furthermore, although we went as far as possible to avoid telegraphing the experimental design and research questions, it could be that the type of person who was willing to participate in our experiment had different treatment effects than the average Facebook user. Third, many of our outcome variables are self-reported, which carries standard problems with interpretation, as well as concerns around experimenter demand effects. We revisit each of these concerns below.

The causal impacts of social media have been of great interest to researchers in economics, psychology, and other fields. Appendix Table A1 lists the 11 previous randomized impact evaluations

that we are aware of.[2] Our deactivation period is many times longer and our sample size an order of magnitude larger than the most relevant prior work. 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 existing literature's sample sizes and duration of intervention.[3]

Our estimates also speak to the literature on measuring the economic gains from information technologies such as search and media, given that companies such as Google, Facebook, Wikipedia, and others provide these services for free.[4] In standard models with consumers who correctly optimize their allocation of time and money, researchers can use time use or monetary valuations of these services to approximate consumer surplus gains, as in Brynjolfsson and Oh (2012) and Brynjolfsson, Eggers, and Gannamaneni (2017). But if social media are in fact harmful and addictive in ways that users do not fully appreciate, then such approaches could overstate consumer surplus gains.

There are two prior studies that shed light on consumers' valuations of Facebook. Brynjolffson, Eggers, and Gannameneni (2017) elicit WTA to abstain from Facebook for one month.[5] Mosquera et al. (2018) elicit WTA to participate in their experiment, which involved a 50 percent chance of Facebook deactivation plus completing a survey, and a 50 percent chance of only completing a survey. Additional assumptions on the perceived cost of survey completion would be required to back out participants' valuation of Facebook access. Our quantitative estimates of average willingness-to-accept to abstain from Facebook align well with these studies. In contrast to both studies, however, we also elicit both the predicted and actual change in Facebook valuation after deactivation, which allows us to quantify how people might mispredict the consumer surplus they derive from Facebook.[6]

[2]The most similar of these papers is Mosquera et al. (2018). Our pre-analysis plan, which we submitted before their paper was released, parallels many of their decisions. Table A1 summarizes the differences between the two papers.

[3]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, specific media providers such as Fox News, and content such as political advertising (Bartels 1993, Besley & Burgess (2001), DellaVigna and Kaplan 2007, Enikolopov, Petrova, and Zhuravskaya 2011, Gentzkow 2006; Gerber and Green 2000, Gerber, Gimpel, Green, and Shaw 2011, Gerber, Karlan & Bergan (2009), Huber and Arceneaux 2007, Martin and Yurukoglu 2017, Olken (2009), and Spenkuch and Toniatti 2016; for reviews, see DellaVigna & Gentzkow (2010), Napoli (2014), Stromberg (2015), Enikolopov & Petrova (2015), and DellaVigna & La Ferrara (2015).

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

[5]The 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." The abstention was generally not enforced: of the ten subjects randomly selected for enforcement, one gave permission. By contrast, our design did not require additional permissions to monitor abstention other than the information we had already collected in the baseline survey, and more than nine in ten participants complied.

[6]Our measurement of state dependence and misprediction connects us to the literature on habit formation and projection bias, including Becker, Grossman, and Murphy (1991), Busse et al. (2014), Conlin, O'Donoghue, and

Sections II and III detail the experimental design and empirical strategy. Section IV presents the impact evaluation, and Section V presents measurements of how consumers value Facebook.

II Experimental Design

II.A Experiment Overview

Table 1 presents the experiment timeline and sample sizes. We timed the experiment so that the main period of Facebook deactivation would end shortly after the 2018 U.S. midterm elections, which took place on November 6th. The experiment has eight parts: recruitment, pre-screen, baseline, midline, endline, and post-endline surveys, post-endline emails, and daily text messages.

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 U.S. 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 people consented to participate in the study.

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

3,234 people finished the baseline survey. Of those, 183 were dropped from the experimental

Vogelsang (2007), Fujiwara, Meng, and Vogl (2016), Gruber and Köszegi (2001), Hussam et al. (2016), and Simonsohn (2010). .

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 and were randomized.

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 we 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 instead of 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 four weeks. 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."

On November 8th, two days after the midterm election, we sent an email invitation to the endline survey. That survey first measured the same outcome variables as in the baseline. All questions were identical, with several exceptions described in Section II.C 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 elicited WTA to stay deactivated for the next four weeks 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, 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.

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

[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 decided not to tell participants the distribution from which the offer prices would be drawn in order to avoid truncating the distribution of elicited WTA.

Settings app, and we asked other participants to just 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 aspects of subjective well-being that were best measured daily instead of in a retrospective endline survey. 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 see when they browse to the user's profile page, the Profile URL returns an error message within the Facebook website stating "Sorry, this content isn't available right now" if and only if the account has been deactivated.[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 the 28 days after endline. Between midline and endline, we verified deactivation approximately four times per day for the Treatment group and once every four days for the Control group. 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.

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 valuations in the BDM.

[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 find their default numeric URLs, which allowed us to correctly measure deactivation even if they changed their customized URL.

II.B 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: Treatment (approximately 35 percent of the sample, with offer price for post-midline deactivation $p = \$102$), Control (approximately 65 percent of the sample, with $p = \$0$), and a third group (approximately 0.2 percent of the sample, with p drawn from a uniform distribution on $[\$0, \$170]$).[9] Treatment and Control were balanced 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 are identified in the sample of participants who were allocated to Treatment or Control and are willing to accept less than \$102 to deactivate Facebook for the four weeks after midline. We call this the "impact evaluation sample."

For the post-endline deactivation, 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 receive offer $p' = 0$ for post-endline deactivation. The post-endline BDM offer price was balanced between Treatment and Control, so two Treatment group members and four Control group members were assigned to positive post-endline offers $p' \in [0, 170]$.

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

II.C Outcome Variables

For the impact evaluation, we consider the outcome variables in the nine families described below. We further collect these families into three groups: "substitution" (Sections II.C.1 and II.C.2), "news and political outcomes" (Sections II.C.3-II.C.5), and "well-being outcomes" (Sections II.C.6-II.C.9). Appendix C presents details on 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 in control group at endline, as described in Anderson (2008). In constructing these indices, we orient the outcome variables so that more 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 "× (-1) in Appendix C.

[9] \$170 was chosen because it was the maximum that we could pay participants without requiring tax-related paperwork.

II.C.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 ask parallel questions about free time activities, but phrased to elicit *changes* in how much time they spent on these things in the past four weeks (between endline and midline) relative to what is normal for them. We phrased the questions in this way in order to more precisely detect changes in self-reported time use caused by the deactivation.

II.C.2 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 Center (2018). At endline, we asked parallel questions, but again phrased to ask about *changes* in how often they got news from these sources over the past four weeks. For the participants who reported having a Twitter handle, we gathered data on number of tweets for the four weeks before baseline began and the four weeks between midline and endline. This allows a non-self-reported measure of one kind of potential substitution away from Facebook.

II.C.3 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 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. 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 two sources, such as "President Trump spoke at the funeral of former Arizona Senator John McCain, honoring the late McCain's wish." 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.

II.C.4 Political engagement

We have two measures of political engagement. First, we measure whether participants voted in the 2018 midterm election. In this draft, we use voting self-reports from the endline survey. In a future draft, we will match participants on name, birth year, and zip code to a voting database supplied to Stanford by L2, a voting data provider. Second, we measure whether participants clicked on any of the links in the post-endline politics email.

II.C.5 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 her own political party minus the number of times she saw news that helped her understand the other party.

Fifth, we asked nine opinions about current political issues, such as "To what extent do you think that free trade agreements between the U.S. 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.[10] Finally, *vote polarization* measures the strength of

[10]Specifically, for each issue or belief question q, we normalize responses to have a standard deviation of one, determine Democrats' and Republicans' average responses μ_q^D and μ_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 response to question q, multiplied by -1 if i is a Democrat. $\tilde{y}_{iq}$

preferences for the congressional candidate of the respondent's party in the midterm election.[11]

II.C.6 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 uses 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 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* 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 (1987) 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)?"

---

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 σ_q as the 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 $Y_i = \sum_q \tilde{y}_{iq}\sigma_q$.

[11] 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.

II.C.7 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. Specifically, the survey question 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 the indicator for whether the respondent interacted with someone who voted the opposite way in the last presidential election plus the indicator for whether the respondent interacted with someone from another country in the last week.

II.C.8 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.

II.C.9 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, *post-endline mobile use* is the natural log of one plus the number of minutes that the participant reported using Facebook on their phone in the post-endline survey.

II.C.10 Secondary outcomes

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

First, we consider the standard generic ballot question: "If the elections for U.S. 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 gathered contributions to political campaigns from the Federal Election Commission 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.

III Empirical Strategy

III.A Sample Demographics, Response Rates, and Compliance with Deactivation

Table 1 shows that our sample is a highly selected group: while almost 1.7 million Facebook users were shown our ads, only 2,846 were randomized into Treatment or Control. Of course, any sample of lab study participants, college students, or online panelists will have similarly large selectivity, even if it cannot be quantified so precisely. How representative is our sample on observable characteristics? Table 2 compares the demographics of our impact evaluation sample to our estimate of the average demographics of adult Facebook users and to the U.S. 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.[12]

Table 3 shows that Treatment and Control are balanced on observables. Table 4 and Figure 2 show that Treatment and Control have statistically equal response rates to the endline and post-endline surveys and subjective well-being text messages.

Table 4 documents remarkably high response rates. Of the 582 Treatment group participants with post-midline WTA less than $102, only seven failed to complete the endline survey. Of the 1082 Control group participants with post-midline WTA less than $102, only 17 failed to complete endline. The average participant responded to 92 percent of daily text messages, well above the 75 percent we had required in order to receive the completion payment.

Figure 3 presents the share of our deactivation checks in which the Treatment and Control

[12]In Appendix Figures A15, A16, A17, and A18, 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. This suggests that re-weighting the sample for representativeness on these observables should not improve generalizability.

groups (with WTA less than $102) were deactivated, by day of the experiment. 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. 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 twice per day. After endline, we returned to checking all participants once per day.

As described above, if Treatment group members with WTA less than $102 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 see two classes of reasons why Treatment group members with WTA less than $102 might not have complied with their assigned deactivation: unintentional and intentional. In some cases, participants told us that they did not know their account was active. One common reason for unintentional re-activations is that Facebook automatically re-activates users accounts when they log in to *another* app or online service (such as Spotify) using their Facebook account credentials. In some other cases, participants said that they had changed their minds and did not want to deactivate.

Appendix Figure A5 shows that the vast majority of Treatment group non-compliance is concentrated among a smaller group of people whose accounts are active most of the time—only a small amount of the total non-compliance is caused by a larger group of people who reactivate for one or two days. Appendix Figure A6 shows that that smaller group of regular non-compliers mostly reports that they know they have logged in to the Facebook app. Because most of Treatment group non-compliance involves people using Facebook, and because Figure 3 demonstrates that a small share of the Control group also remains deactivated, we will estimate treatment-on-the-treated effects using instrumental variables.

III.B Pre-Analysis Plan

We submitted our pre-analysis plan on October 12, immediately after the midline survey was completed. We submitted a slightly updated PAP 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 outcomes are "secondary." (Figures 3, 4, 5, 7, and 11 appear as figure shells in the pre-analysis plan.) Second, it 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.

III.C 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.[13] 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 ν_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), τ 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 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 almost exactly identical. Therefore, we focus on Equation (1) because it is slightly simpler. Appendix D.B presents results using Equation (2).

The sample for all regressions is the "impact evaluation sample"—participants who were allocated to Treatment or Control and were willing to accept less than $102 to deactivate Facebook for the four weeks after midline.

Our figures present point estimates and unadjusted 95 percent confidence intervals from estimates of Equations (1) and (2). We do formally discuss p-values, although readers should bear in mind that any implied Type I error rates are not adjusted for multiple hypothesis testing.

[13] $\boldsymbol{Y}_i^b$ 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.

IV Impact Evaluation

This section presents the treatment effects of Facebook deactivation. The next three subsections present estimates for each of our three outcomes groups: substitution, news and political outcomes, and well-being outcomes. We then present heterogeneous treatment effects. Finally, we discuss robustness checks and effects on secondary outcomes.

IV.A Substitutes for Facebook

Figure 4 presents treatment effects on substitutes for Facebook: substitute news sources and substitute time uses.

The top group of outcomes measures self-reported time use. The local average treatment effect of deactivation was to reduce Facebook minutes by about 1.6 standard deviations. Appendix Table A3 presents estimates of effects without normalizing into standard deviation units; the LATE is a 59.6 minute reduction in daily average Facebook use.

Facebook deactivation also reduced time on non-Facebook social media use and non-social media online activities. While one might naturally think that these activities are the closest substitutes for Facebook, our qualitative interviews and open-answer survey questions suggest two reasons for complementarity. First, many people use their Facebook credentials to log in to other apps and websites, and we informed participants that Facebook would automatically reactivate their account if they logged in to these other websites. Second, deactivating Facebook reduces the appeal of using smartphones and web browsers in the first place, and if people pick up their phones less overall, they will not use apps on their phones. The Treatment group reported that they spent their newly available time to television and other solitary activities as well as seeing friends and family. The "substitute time uses index" variable, which does not include Facebook minutes, shows an increase in overall non-Facebook activities.

The bottom group of outcomes in Figure 4 measures self-reported news sources. As one might expect, 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, and the point estimates for print, radio, and TV news are all positive, although not statistically significant. Facebook deactivation has a positive but insignificant effect on Twitter use.

Overall, these results suggest that Facebook crowds out offline activities and crowds in other online activities, including other social media.

IV.B Effects on News and Political Outcomes

Figure 5 presents treatment effects on news and political outcomes: news knowledge, political engagement, and political polarization.

Within the news knowledge group, the top three rows of the figure show that deactivation reduced how much people reported they followed news about politics and about president Trump, as well as the average minutes per day spent consuming news. Appendix Table A3 shows that deactivation reduces news consumption by tk minutes per day. In addition to self-reported news consumption, deactivation also reduces news knowledge, as measured by our news knowledge quiz, although there is no statistically detectable effect on fake news knowledge. Overall, deactivation reduces the news knowledge index by about 0.2 standard deviations.

There are no statistically detectable effects on political engagement.

Deactivation affects several measures of political polarization and exposure to polarizing content. Deactivation increased other party understanding—i.e., the number of times that participants reported seeing news that made them better understand the point of view of the other political party relative to their own political party. Deactivation also increases issue polarization. Overall, deactivation reduces the political polarization index by about 0.13 standard deviations.

Figure 6 illustrates the change in issue polarization between baseline and endline. For each issue position question, we normalize it to have a standard deviation of one and a mean equal to the average of Democrats' average answer and Republicans' average answer, and we re-sign (if needed) so that Republicans' average answer is more positive. We then sum each individual's normalized answers across the nine issues, weighting issues by the standard deviation of within-person endline - baseline differences in order to place more weight on questions where opinions are more malleable over the deactivation period. For the figure, we then plot the distribution of individual issue opinions for Republicans and Democrats. The figure shows that Treatment group Democrats and Republicans move visible toward the center relative to the Control group.

Overall, these results suggest that Facebook plays an important role in helping people stay informed about current events. However, the information that people receive appears to favor their own party, and thus increases polarization of views on political issues.

IV.C Effects on Well-Being Outcomes

Figure 7 presents estimates on what we broadly call well-being outcomes: subjective well-being, social interactions, opinions about Facebook, and post-experiment Facebook usage.

The top group of outcomes is subjective well-being (SWB). These are re-signed so that more positive represents better SWB—for example, the "depressed" variable is multiplied by (-1). We see no statistically significant changes on any of the ten SWB measures, and we can rule out with 95 percent confidence that deactivation increases our SWB index by more than 0.13 standard deviations.[14]

[14] Appendix Figure A20 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.

This non-result contrasts sharply with results of correlation studies cited in the introduction that suggest that social media use decreases SWB. Many of these studies have specific designs and outcomes that don't map closely to our paper, so it is difficult to directly compare effect sizes. For comparison to non-experimental results, 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. The baseline correlation between our SWB index and Facebook use is about five times larger than the experimental estimate of the treatment effect of deactivation (about 0.22 standard deviations compared to about 0.04 standard deviations), and the point estimates are highly statistically significantly different. Controlling for basic demographics brings down the non-experimental estimate to about 0.15 standard deviations, but this is still economically and statistically larger than our experimental estimate. Appendix Figure A21 presents the full results for all SWB outcomes.[15]

One possible explanation for the difference between the experimental and non-experimental estimates is that social media use takes a longer period to affect our measures of subjective well-being. If this is true, the experimental estimates would be much larger after a longer deactivation period, and the non-experimental estimates could plausibly be unbiased estimates of long-run effects. This explanation might be more plausible for some outcomes (such as life satisfaction over the past four weeks), and less plausible for other outcomes (such as the momentary assessments of happiness and emotions measured by our daily text messages). A second explanation for the difference between the experimental and non-experimental results is simply that unobservables or reverse causality confound the relationship between social media use and SWB—for example, does social media use cause loneliness, or does loneliness cause social media use? This uncertainty highlights the importance of our experimental estimates.

Deactivation increases the count of offline activities that people reported doing at least once last week by about 0.1 standard deviations. Notwithstanding, 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 presential election or interacted with someone from another country). Overall, there are no effects on the social interaction index.

The third group of outcomes in Figure 7 measure participants' opinions about Facebook. These are re-signed so that more positive means more positive opinions—for example, agreement with the statement that "deactivation was good for me" is multiplied by (-1). The results are mixed. Deactivation increases the extent to which people think Facebook helps them follow the news better, and it also makes people more likely to say that people would miss Facebook if they stopped using

[15]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.

it. On the other hand, people who deactivated for four weeks instead of 24 hours were more likely to say that their deactivation was good for them. Deactivation increases both the positive impacts and the negative impacts variables, i.e. it makes people more aware of (and willing to write more about) both positive and negative aspects of Facebook. The point estimate on positive impacts is slightly larger in absolute value. Overall, deactivation increases the Facebook opinions index.

The final group of outcomes in Figure 7 measures planned and actual post-experiment use. Deactivation causes people to plan to use Facebook less, to be more likely to click on our email offering information about ways to limit social media usage, and to re-activate much more slowly (as also illustrated in Figure 3). Our point estimates suggest that deactivation also reduced use of the Facebook mobile app as measured in the post-endline survey in early December. Overall, deactivation appears to decrease post-experiment use. These results are consistent with the idea that Facebook is habit-forming.

Focusing on treatment effects risks obscuring information on the distribution of responses. Figure 8 presents the distribution of Treatment and Control endline responses to four key questions that illustrate people's opinions about Facebook and their hopes for future usage. Both Treatment and Control tend to agree that Facebook helps to follow the news, and deactivation strengthens that view. 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 *weakens* that view. 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. Finally, both Treatment and Control plan to use Facebook less after the experiment, and deactivation makes people more likely to plan to reduce usage.

IV.D 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. Perhaps surprisingly, the effects are not larger for people who report being heavier users. 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 and political polarization are larger for people who get more news from Facebook. The pre-analysis plan specified that we would limit these tests to only the news and political outcomes in Section IV.B,

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," which was

suggested by Facebook's well-being team as a key measure of the type of usage that they think improves well-being. Active vs. passive users are defined as having above- vs. below-median sum of their two responses to these questions. We were interested in this moderator because of a set of papers 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 effects of deactivation on the subjective well-being index. If anything, deactivation caused passive users to have an *improved* opinion of Facebook. The pre-analysis plan specified that we would limit these tests to the well-being outcomes in Section IV.C.

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 still see zero statistical effects on subjective well-being, even during peak usage times.

V 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.

V.A Model

V.A.1 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 from period to period depending on two factors: *state-dependence* and *misprediction*. State dependence will be a reduced form statistic capturing various mechanisms through which past consumption affects future utility, 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 allowing people to misforecast how past consumption will affect future utility, including projection bias and learning.

We use these "reduced form" concepts in order to map directly to our experimental design. State dependence would imply that the Treatment group values Facebook differently at endline than they did at midline, and they know that this will happen. Misprediction would imply that the Treatment group values Facebook 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 ξ is average utility from Facebook, ω_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 ϕ 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 τ, consumers predict their utility for that period and for future periods. Predicted utility for $t \geq \tau$ is

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

If using Facebook in the most recent period, people overestimate utility from Facebook by amount α.[16] $\alpha > 0$ could result from projection bias, or from not having had recent experience without using Facebook.

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

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

We assume that people foresee that they will always use Facebook in future periods, even if they do not in this period: $f_t = 1, \forall t > \tau$. This is approximately true in our data, and it dramatically simplifies the dynamic problem.

V.A.2 Valuation of Facebook

We can now write the three WTAs elicited in our experiment in terms of the model. We denote $v_{t,\tau}(f_{t-1})$ as the valuation of Facebook for period t as evaluated at the beginning of period τ, given f_{t-1}. People determine $v_{t,\tau}(f_{t-1})$ by the payment that would be required to equalize $\tilde{U}_\tau(f_\tau = 0)$ with $\tilde{U}_\tau(f_\tau = 1)$. We define the four weeks after midline as $t = 0$ and the four weeks after endline as $t = 1$. The WTA for post-midline deactivation is

$$v_{0,0}(1) = \underbrace{\phi + \alpha + \xi + \omega_0}_{\text{post-midline Facebook utility loss}} + \underbrace{\phi}_{\text{post-endline Facebook utility change}} \tag{6}$$

In the model, WTA for post-endline deactivation may differ for Treatment vs. Control, because of different state dependence. This WTA may also depend on whether it is elicited at midline or

[16] The timing maps to the timing of our experiment, where people's valuations for period t are elicited at the end of period $t - 1$.

endline, because of misprediction. We define $\Delta v := v_{1,1}(f_0) - v_{1,0}(f_0)$ as the difference between WTA for post-endline deactivation as of endline and WTA for post-endline deactivation as of midline. Table 5 presents the WTAs for post-endline deactivation for Treatment and Control as of midline and endline. For the Control group, Δv results from the utility shock ω_1. For the Treatment group, Δv results from the utility shock and misprediction: $\omega_1 - \alpha$. The difference in differences is thus $-\alpha$, our measure of misprediction.

V.A.3 Consumer Surplus

In the model, the consumer surplus from Facebook is

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

A standard way of measuring consumer surplus, e.g. in Brynjolffson et al. (2017), is to approximate it with the willingness-to-accept to deactivate over one period, which is v^b as written in Equation (6). But v^b differs from actual consumer surplus by the following ratio:

$$\frac{CS' - CS}{CS'} = \frac{(\phi + \alpha + \omega_0)}{v_{0,0}(1)} \tag{8}$$

In words, WTA differs from consumer surplus for three reasons: state dependence ϕ, misprediction α, and the time effect ω.

V.B Results

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 have valuations greater than \$500. After winsorizing valuations at \$1000, the mean is \$200.

Figure 11 presents the average of WTA in Treatment and Control for the impact evaluation sample (those with $v_{0,0}(1)_i < 102$) in each of the three elicitations: at midline for post-midline deactivation, at midline for post-endline deactivation, and at endline for post-endline deactivation. The Treatment group's post-endline valuation jumps substantially relative to its post-midline valuation. From additional qualitative interviews, we believe that some of this is due to convex costs of deactivation. For example, many people in the Treatment group reported in the post-endline open answer survey that they were much less comfortable deactivating for eight weeks instead of four, as they would have to make much more serious alternative arrangements to communicate with friends, co-workers, and schoolmates over a longer deactivation. However, some of this also appears to have been caused by some combination of anchoring and misunderstanding the incentive-compatibility

of the BDM procedure. [17] Thus, the treatment effect on post-endline WTA relative to post-midline WTA should not be interpreted as a measure of the state dependence term ϕ in our model.

We focus more on the treatment effect on Δv, the difference between WTA for post-endline deactivation as of endline minus the WTA for post-endline deactivation as of midline. The Control group's Δv is slightly positive, which in our model represents a time effect ω_1. In open-answer questions, many people reported that they didn't want to deactivate during the beginning of the holidays, although they may not have foreseen this as of the midline survey on October 11th. By contrast, the Treatment group has a slightly negative Δv. Thus, the treatment effect on Δv is somewhat negative. In our model, this is interpreted as the Treatment group slightly over-predicting their valuation of Facebook.

We can formally estimate the misprediction α using the following regression:

$$\Delta v_i = \alpha D_i + \rho v_{0,0}(1)_i + \nu_s + \varepsilon_i, \tag{9}$$

instrumenting for D_i with T_i. This parallels Equation (1).

Table 6 presents results. Relative to the Control group, the Treatment group reduced its post-endline valuation by an average of \$18, suggesting slight over-prediction equal to approximately 13 percent of their average post-endline valuation.

V.C Caveats

There are a number of caveats to this calculation. First, our sample explicitly excludes people who use Facebook 15 minutes per day or less, and who thus likely value Facebook less. Furthermore, our sample may not be representative on unobservables. Second, we estimate valuation of Facebook holding networks fixed. Due to network externalities, valuations would be quite different for a social network with more or fewer other people, and would thus be quite different if participants' friends had also deactivated, instead of deactivating in isolation. Third, we consider WTA for only four-week periods, and our quantitative and qualitative evidence suggest 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. Finally, while we lean toward interpreting the WTA elicited after endline as a measure of "true" utility, while post-endline WTA elicited as of

[17] As discussed in Section II, the survey instrument correctly explained the BDM procedure, including saying that "the computer has randomly generated an amount of money to offer you," and all participants correctly answered comprehension questions. Notwithstanding, many Treatment group members with valuation less than \$102 told us in the post-endline survey that they reported higher valuations for post-endline deactivation because they were surprised that the experimenters were willing to pay so much and wanted to see if they could earn more money from the experiment.

midline to be subject to mis-prediction, there is certainly some ambiguity. What is clear is that our measurements suggest at least some uncertainty as to whether active consumers' valuations of social media should be taken seriously as measures of consumer surplus.

VI Conclusion

The rise of social media has provoked both optimism about potential benefits to society, and concern about harms ranging from addiction to depression to political polarization. We present a large-scale, randomized evaluation of the welfare impacts of Facebook, focusing on US users in the run-up to the 2018 midterm election. We measure the willingness to accept of 2,844 Facebook users to deactivate their Facebook accounts for a period of four weeks, then randomly assign a subset to actually do so in a way that we verify. Using a suite of outcomes from both surveys and direct measurement, we show that Facebook deactivation (i) reduces online activity including other social media, while increasing offline activities including time with family and friends; (ii) has precisely estimated zero effects on subjective well being; (iii) reduces both factual knowledge of news and political polarization; and (iv) causes a persistent reduction in use of Facebook following the experiment. We combine these facts with the incentivized valuations we elicit to calibrate a simple behavioral model. We conclude that forces like addiction and mis-prediction may lead to some over-consumption, but that the magnitude of these mistakes are likely small relative to the total consumer surplus gains from Facebook.

References

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.

Brynjolfsson, Erik, Felix Eggers, and Avinash Gannamaneni. 2017. "Using Massive Online Choice Experiments to Measure Changes in Well-being." NBER Working Paper.

Csikszentmihalyi, Mihaly and Reed Larson. 1987. "Validity and reliability of the Experience-Sampling Method." *The Journal of Nervous and Mental Disease* .

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

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.

Gentzkow, Matthew. 2016. "Polarization in 2016." Toulouse Network for Information Technology Whitepaper.

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.

Gruber, Jonathan and Botond Köszegi. 2001. "Is Addiction "Rational"? Theory and Evidence." *The Quarterly Journal of Economics* .

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.

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.

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.

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

Stone, Arthur A. and Saul Shiffman. 1994. "Ecological momentary assessment (EMA) in behavorial medicine." *Annals of Behavioral Medicine* 16 (3):199–202.

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

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

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).

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* .

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).

Table 1: **Experiment Timeline and Sample Sizes**

Phase	Date	Sample size
Recruitment and baseline	September 24 - October 3	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	October 11	N=3,015 began midline N=2,844 finished midline, of which: N=1,661 in impact evaluation sample
Endline	November 8	N=2,823 began endline N=2,795 finished endline, of which: N=1,637 in impact evaluation sample
Post-endline	December 3	N=2,137 reported Facebook use, of which: N=1,214 in impact evaluation sample

Table 2: **Sample Demographics**

	(1) Sample population	(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

Notes: Column 1 presents average demographics for the impact evaluation sample: participants who were allocated to Treatment or Control and were willing to accept less than $102 to deactivate Facebook for the four weeks after midline. Column 2 presents our estimate of average demographics of American adults with a Facebook account, using data from tk Pew. Column 3 presents average demographics of American adults.

Table 3: **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)	
Active browsing	0.14	0.16	0.73
	(0.98)	(0.97)	
Facebook minutes	99.58	98.31	0.79
	(92.41)	(90.56)	
Get news from Facebook	3.47	3.43	0.45
	(1.12)	(1.06)	
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 allocated to Treatment or Control and were willing to accept less than $102 to deactivate Facebook for the four weeks after midline. Column 3 presents p-values of tests of differences in means between the two groups.

Table 4: **Survey Response and Treatment Compliance Rates**

Variable	(1) Treatment Mean/SD	(2) Control Mean/SD	T-test P-value (1)-(2)
Completed Endline	0.99	0.98	0.54
	(0.11)	(0.12)	
Share completed text messages	0.92	0.93	0.45
	(0.20)	(0.18)	
Completed Post-Endline	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 allocated to Treatment or Control and were willing to accept less than $102 to deactivate Facebook for the four weeks after midline. Column 3 presents p-values of tests of differences in response rates between the two groups.

Table 5: **Model: Willingness-to-Accept to Deactivate Facebook after Endline**

		Group		
		Treatment	Control	Difference
Time of elicitation	Midline	$\alpha + \xi + \phi$	$\phi^1 + \alpha + \xi + \phi$	$-\phi^1$
	Endline	$\xi + \omega_1 + \phi$	$\phi^1 + \alpha + \xi + \omega_1 + \phi$	
	Difference	$\omega_1 - \alpha$	ω_1	$-\alpha$

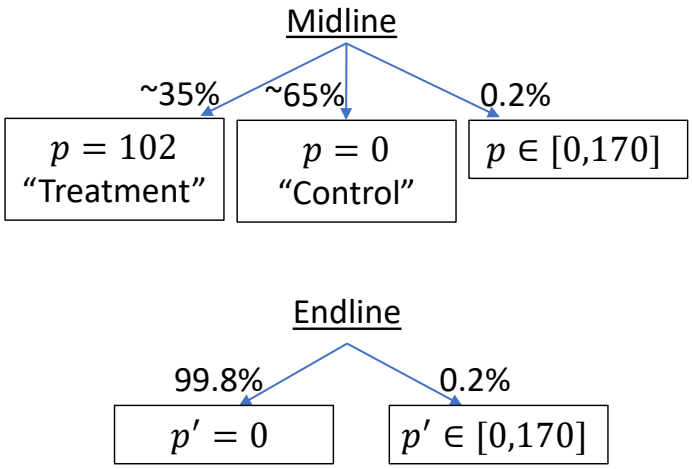
Notes: This table presents the equations for willingness-to-accept to deactivate Facebook after the endline survey, for Treatment and Control, elicited at midline and endline, in the context of the model laid out in Section V. The bottom row presents the endline - midline difference, and the right column presents the Treatment - Control difference. The bottom right cell presents the difference-in-differences.

Table 6: **Misprediction: Treatment Effects on the Post-Endline WTA Update**

	(1)
Share of time deactivated	-18.22**
	(7.73)
Observations	1,634

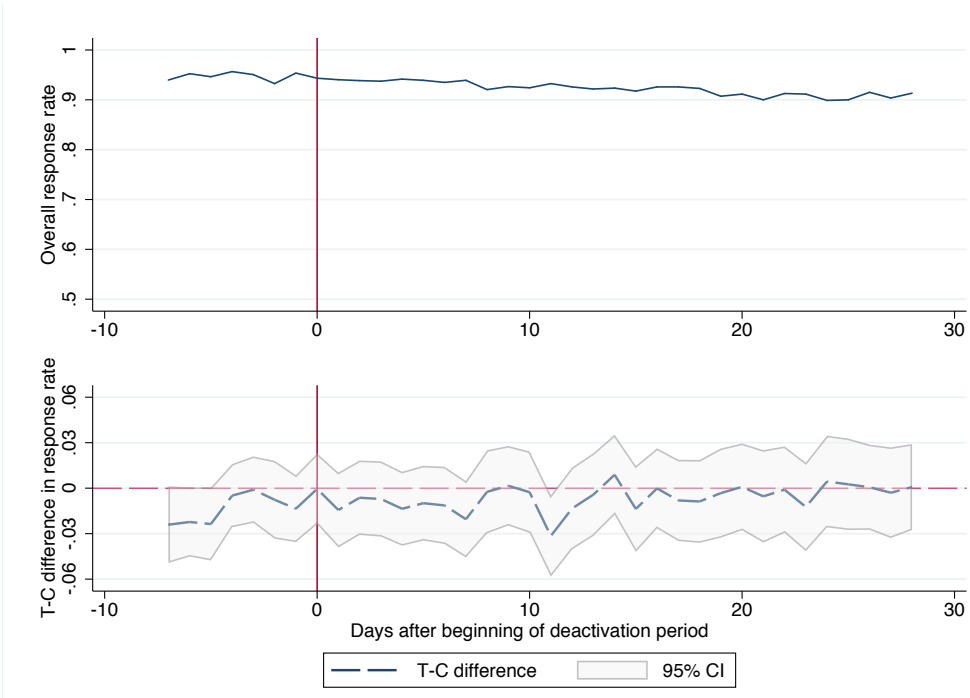
Notes: This table presents estimates of Equation (9). Standard errors are in parentheses. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Figure 1: **Experimental Design**



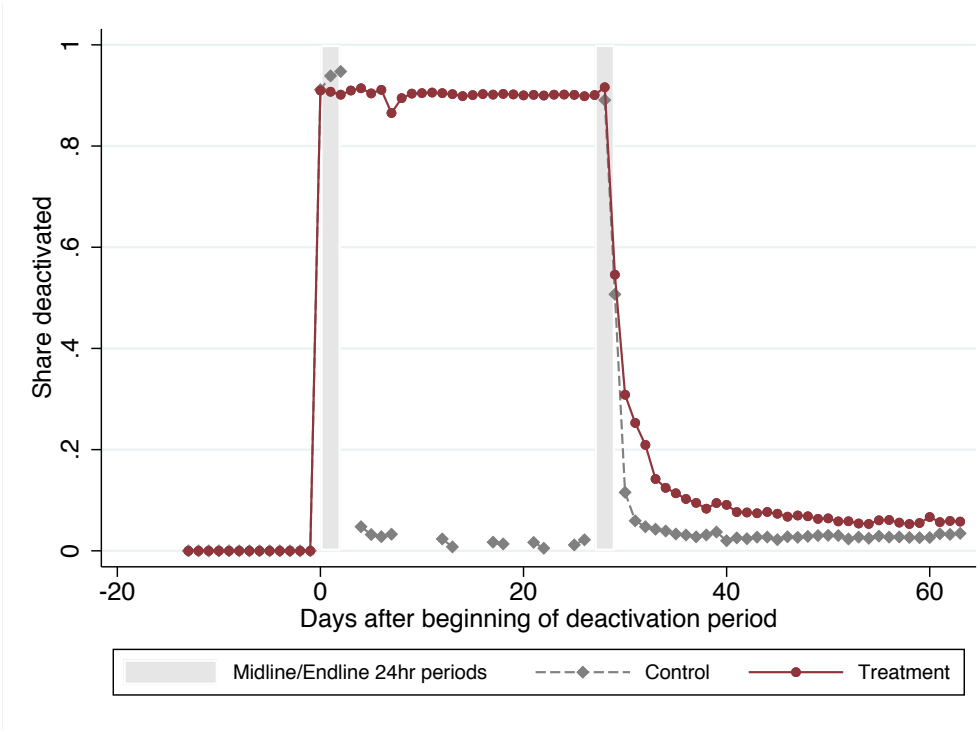
Notes: This figure presents the share of the sample assigned to each price p offered for post-midline deactivation and to each price p' offered for post-endline deactivation.

Figure 2: **Response Rates to Daily Text Messages**



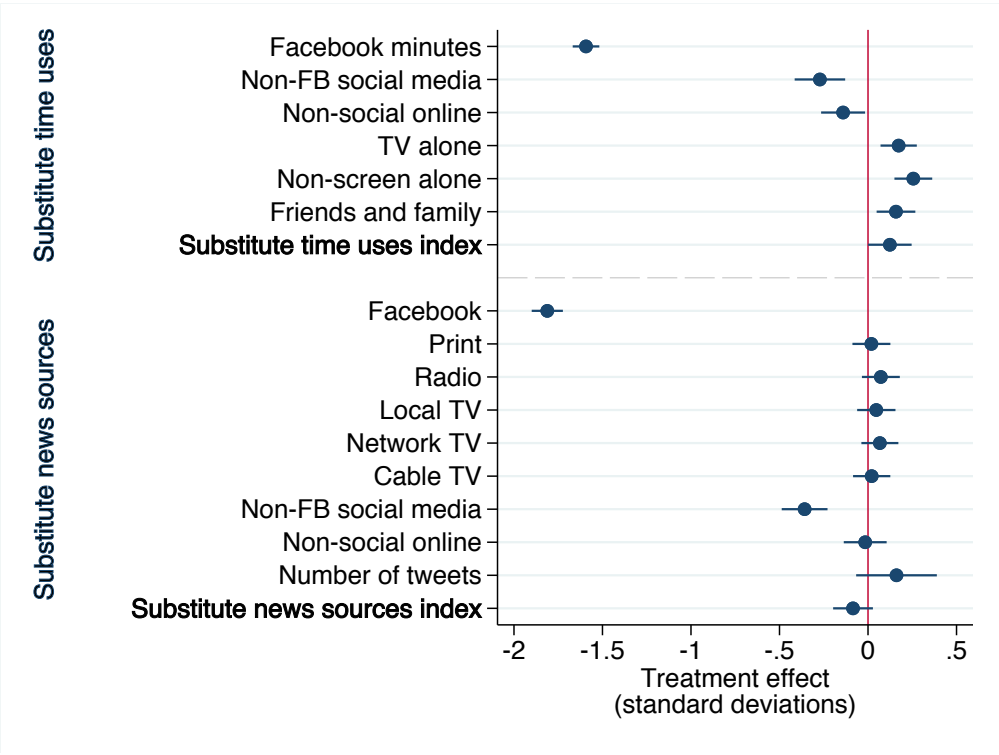
Notes: The figure shows response rates to the SMS survey and the difference in response rates between Treatment and Control, for the impact evaluation sample: participants who were allocated to Treatment or Control and were willing to accept less than $102 to deactivate Facebook for the four weeks after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

Figure 3: **Probability of Being Deactivated**



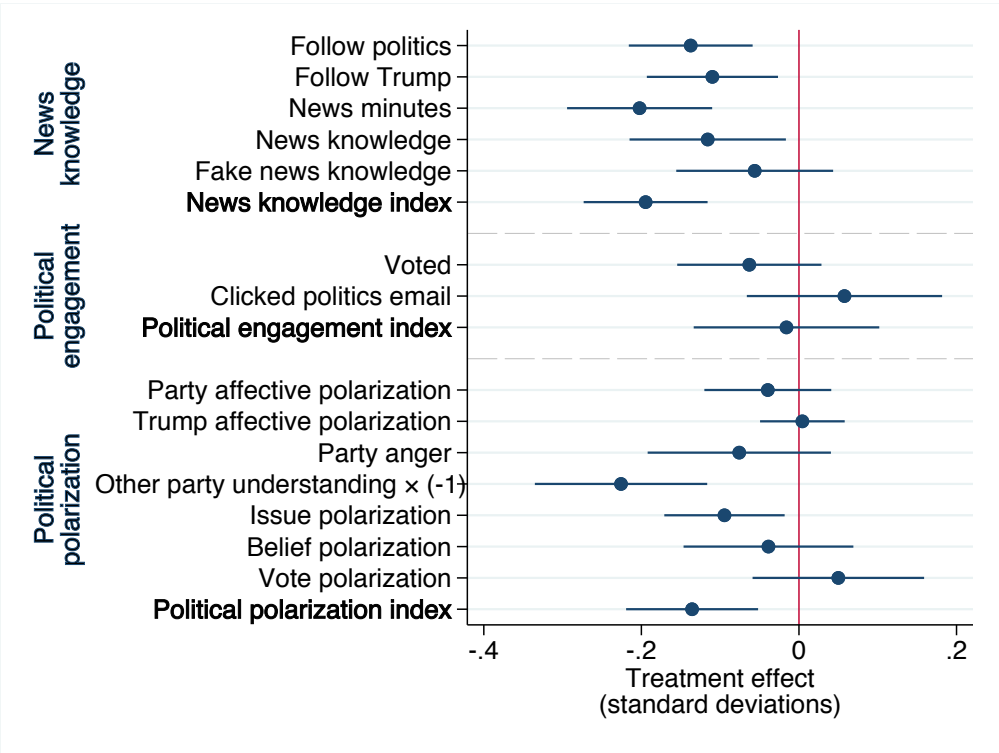
Notes: This figure shows the share of the Treatment and Control groups that had their Facebook accounts deactivated, by day of the experiment, for the impact evaluation sample: participants who were allocated to Treatment or Control and were willing to accept less than $102 to deactivate Facebook for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

Figure 4: **Substitutes for Facebook**



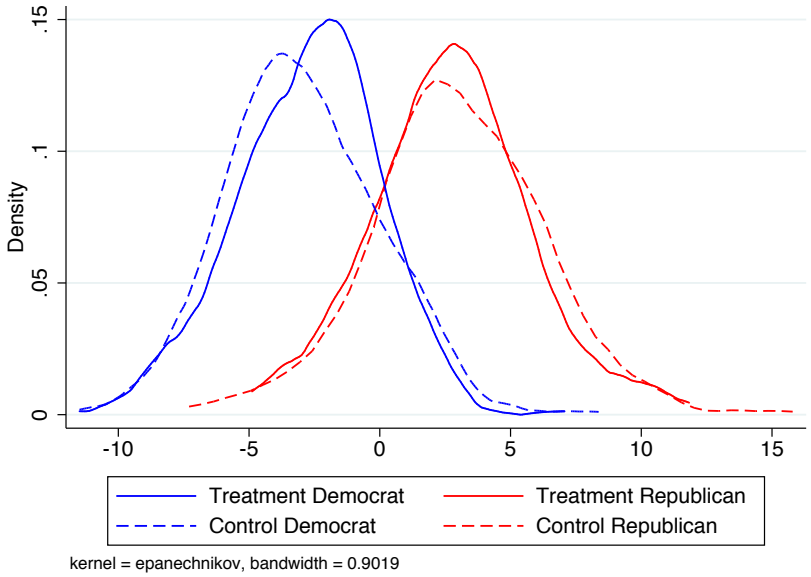
Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section II.C for variable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

Figure 5: **Effects on News and Political Outcomes**



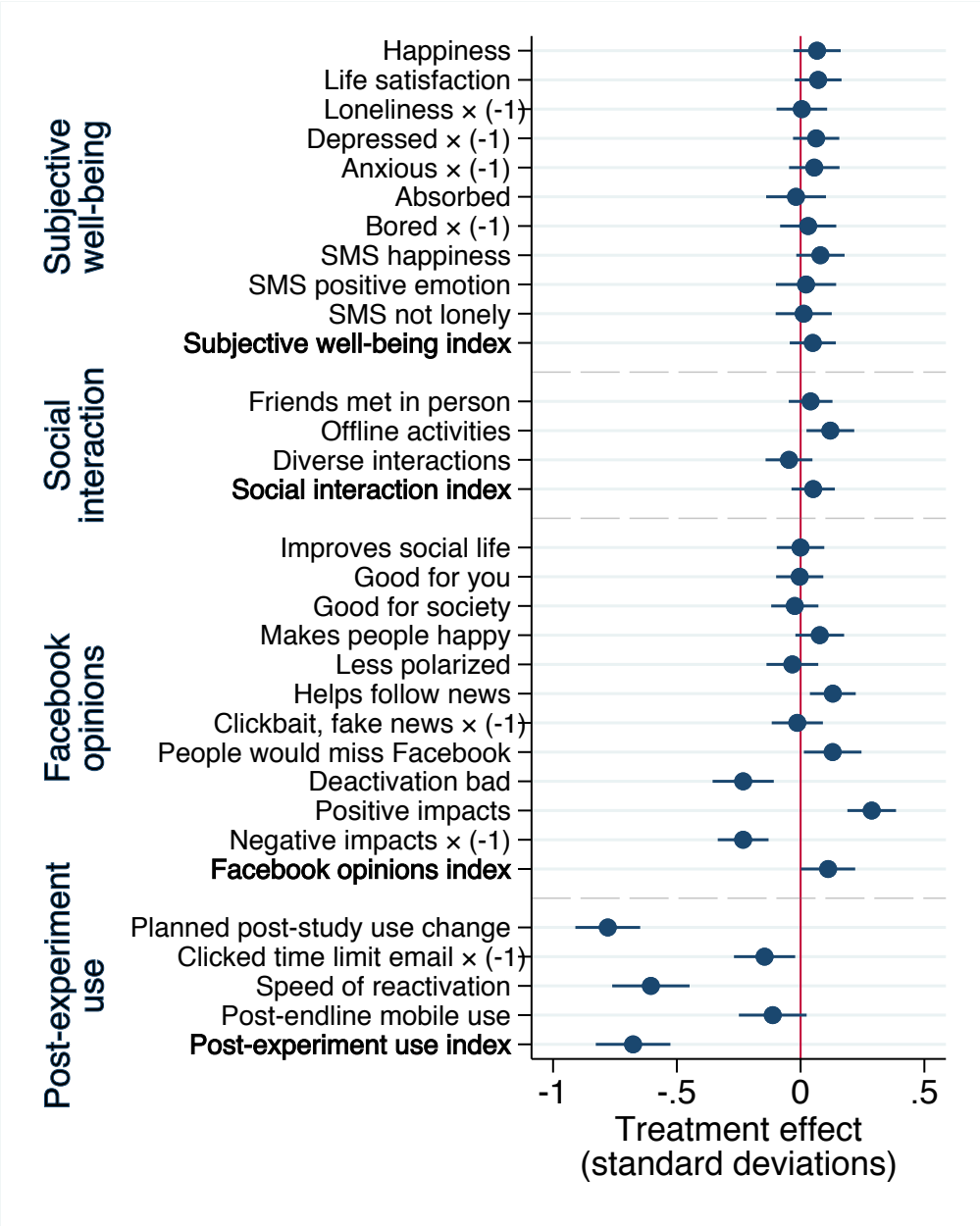
Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section II.C for variable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

Figure 6: **Issue Opinions by Party at Baseline and Endline**



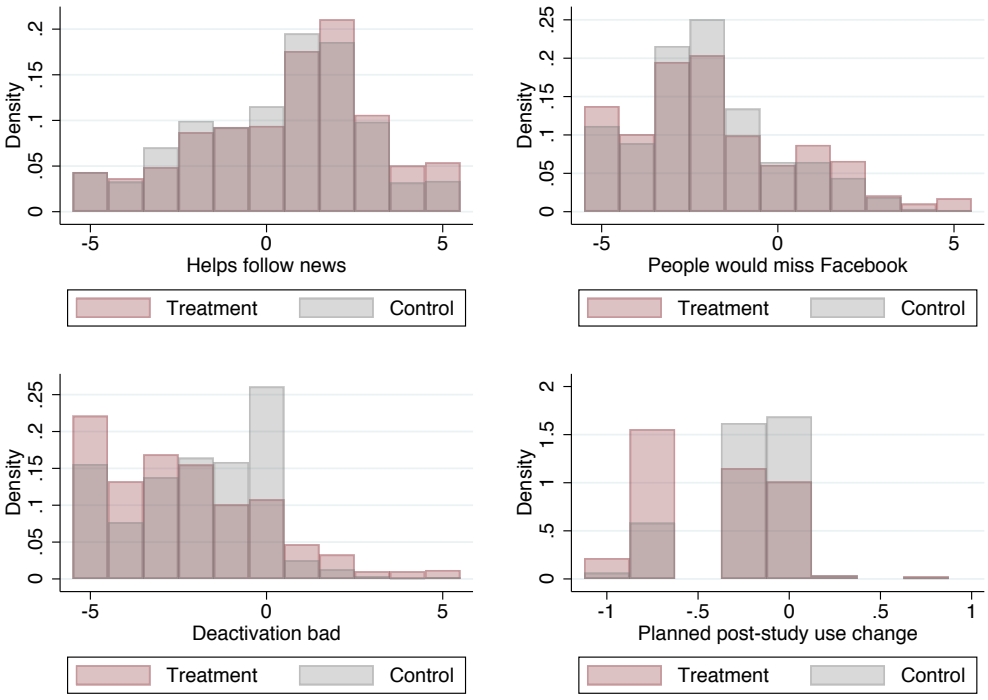
Notes: This figure presents kernel density plots of issue opinions for Democrats and Republicans at baseline and endline. Issue opinions are attitudes about nine current political issues on a scale from -5 to +5, such as "To what extent do you think that free trade agreements between the U.S. and other countries have been a good thing or a bad thing for the United States." See Appendix C for a list of all nine issues questions. We re-sign responses so that the average Republican response is more positive than the average Democratic response, and then sum across questions to construct each individual's issue opinion index. See footnote 10 for specifics.

Figure 7: **Effects on Subjective Well-Being, Social Interactions, and Opinions about Facebook**



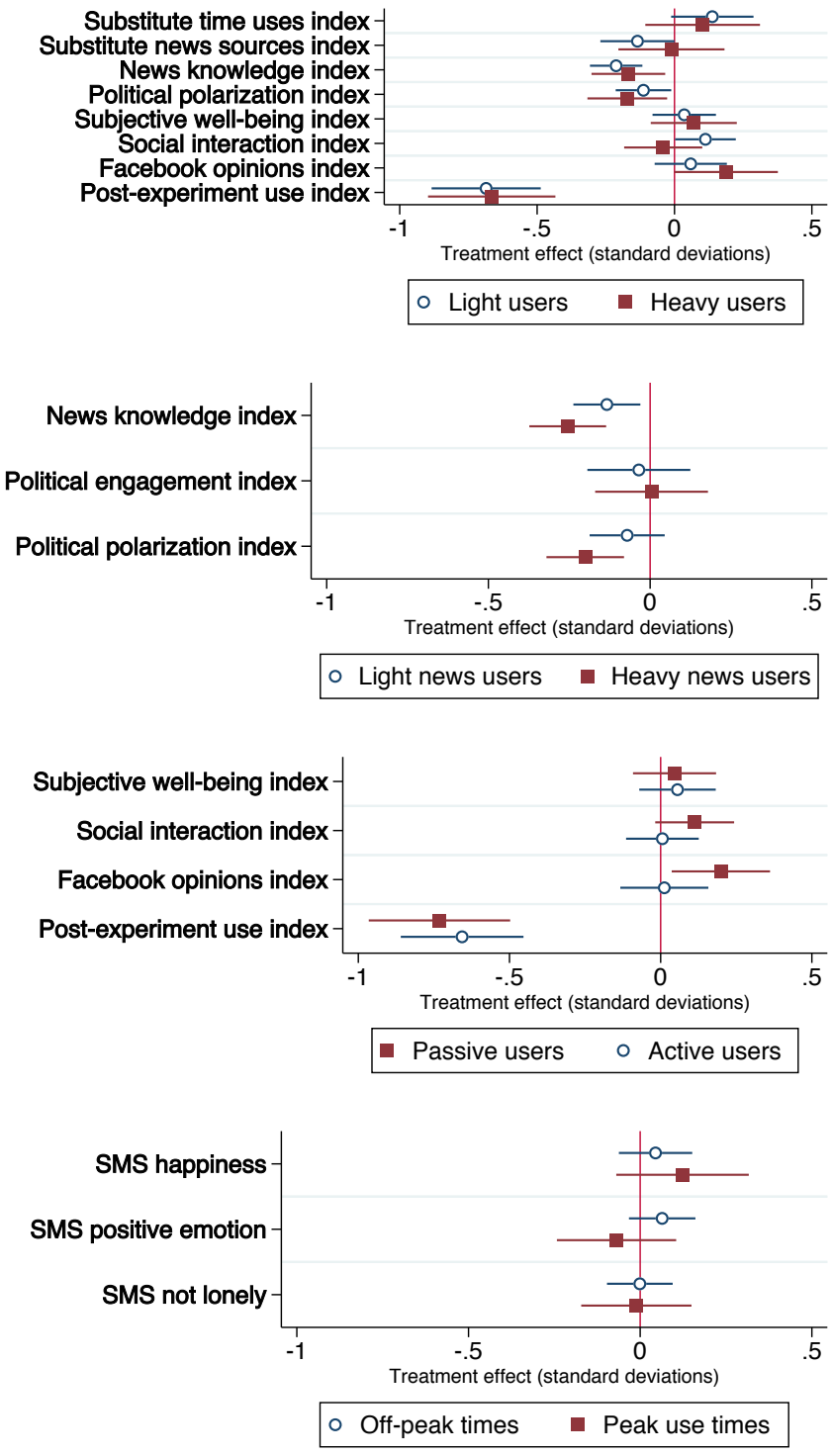
Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1). All variables are normalized so that the control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section II.C for variable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

Figure 8: **Effects on Subjective Well-Being, Social Interactions, and Opinions about Facebook**



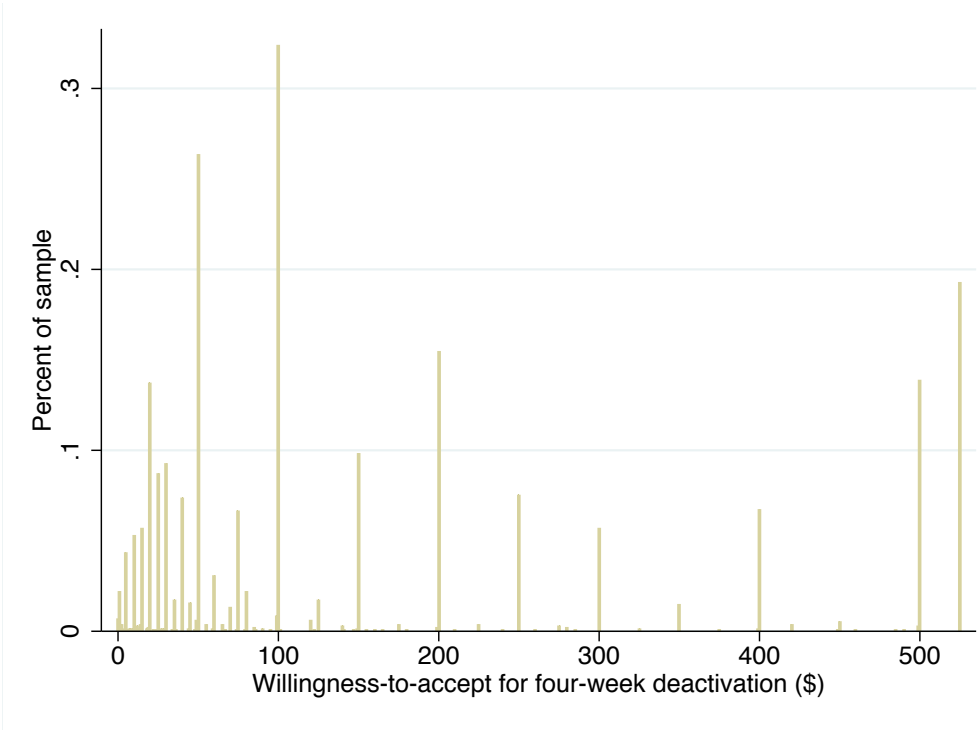
Notes: This figure presents the distribution of responses in Treatment and Control for four questions on the endline survey. See Section II.C for variable definitions.

Figure 9: **Heterogeneous Treatment Effects**



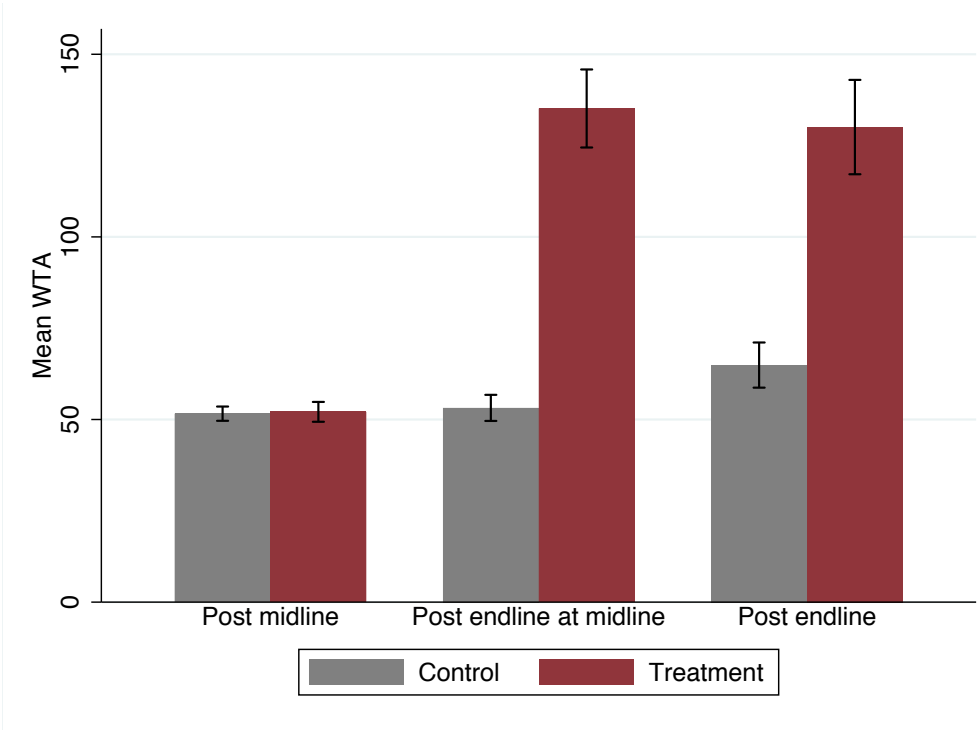
Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1), for subgroups defined by the primary moderators in our pre-analysis plan. All variables are normalized so that the control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section II.C 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 allocated to Treatment or Control and were willing to accept less than $102 to deactivate Facebook for the four weeks after midline. The first pair of bars is the mean WTA for post-midline deactivation. The second pair of bars is mean WTA for post-endline deactivation, as elicited in the midline survey. The third pair of bars is mean WTA for post-endline deactivation, 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

A Related Literature: Facebook Impact Evaluations

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

Paper	N	Population	Intervention	Length	Enforcement	Outcomes	PAP
Gonzales & Hancock (2011)	63	College	Look at profile vs. see self in mirror	3 minutes	None	Self-esteem	No
Deters & Mehl (2012)	86	College	Post more status updates	1 week	Friend request, scrape profile*	SWB	No
Mabe, Forney & Keel (2014)	84	College women	Browse Facebook vs. research ocelots	20 minutes	None	Eating disorder risk	No
Sagioglou & Greitemeyer (2014)	263	MTurk	Browse Facebook	20 minutes	None	SWB	No
Fardouly et al. (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 & Lowe (2016)	197	Greek, without accounts	Sign up	6 months	Payment sent to Facebook account*	Voting, civic engagement	No
Trumholt (2016)	886**	Danish	Not log in	1 week	Self-report	SWB	No
Marotta & 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 FB, Snapchat, Instagram to 10 minutes/day	4 weeks	Weekly screenshots of app time use	SWB	No
Mosquera et al. (2018)	151	College	Not log in	1 week	Check "last active"	News, SWB, WTA for experiment***	No
Allcott, Braghieri, Eichmeyer & Gentzkow (2018)	1642	U.S. Facebook ads	Deactivate	4 weeks	Check URLs	News, voting, polarization, SWB, WTA, WTA changes	Yes

* 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.

B Experimental Design Appendix

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



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



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.

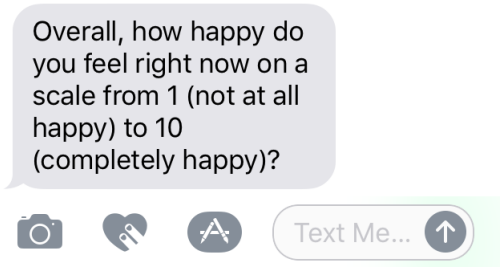
Donate to the Republican Party to help prepare for future elections

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

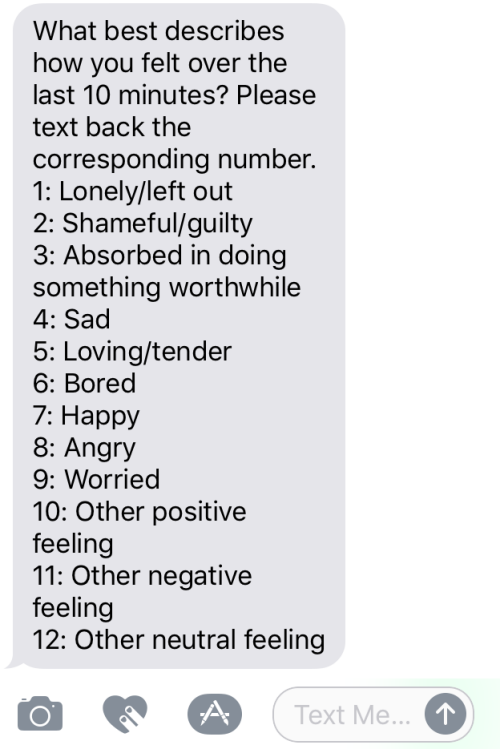
The Stanford Online Experience Research Study Research Team

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

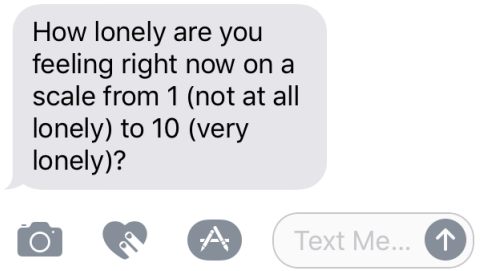
(a) **Happiness**



(b) **Primary Emotion**



(c) **Loneliness**



C 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)*
	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]
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)?
	Substitute news sources
	Over the past four weeks, how often did you... [Never, Hardly Ever, Sometimes, Fairly Often, Very Often]
Facebook	...get news from Facebook *(not included in substitute news sources index)*
Print	...read any newspapers in print?
Radio	...listen to the news on the radio?
Local TV	...watch local television news?
Network TV	...watch national evening network television news (such as ABC World News, CBS Evening News, or NBC Nightly News)?
Cable TV	...watch cable television news (such as CNN, the Fox News cable channel, or MSNBC)?
Non-FB social media	...get news from social media sites other than Facebook (e.g. Twitter or Snapchat)?
Non-social online	...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)	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.
	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)	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 U.S. 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.
	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 U.S. 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 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.) List as many recent (last 4 weeks) news events you can think of that made you angry at the 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] 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 U.S. and other countries have been a good thing or a bad thing for the United States? Overall, would you say that blacks or whites are treated more fairly in dealing with the police?

	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?
	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?
	On the whole, do you think immigration is a good thing or a bad thing for this country today?
	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?
	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?
	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?
	To what extent do you think President Trump is honest and trustworthy?
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 $\times$ (-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)?
	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 (",").
	Which of the following activities did you do at least once last week? Check all that apply
Offline activities	Go out for dinner
	Go to the cinema
	Talk to friends on the phone
	Go to a party
	Get together with friends
	Go to a shopping mall
	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

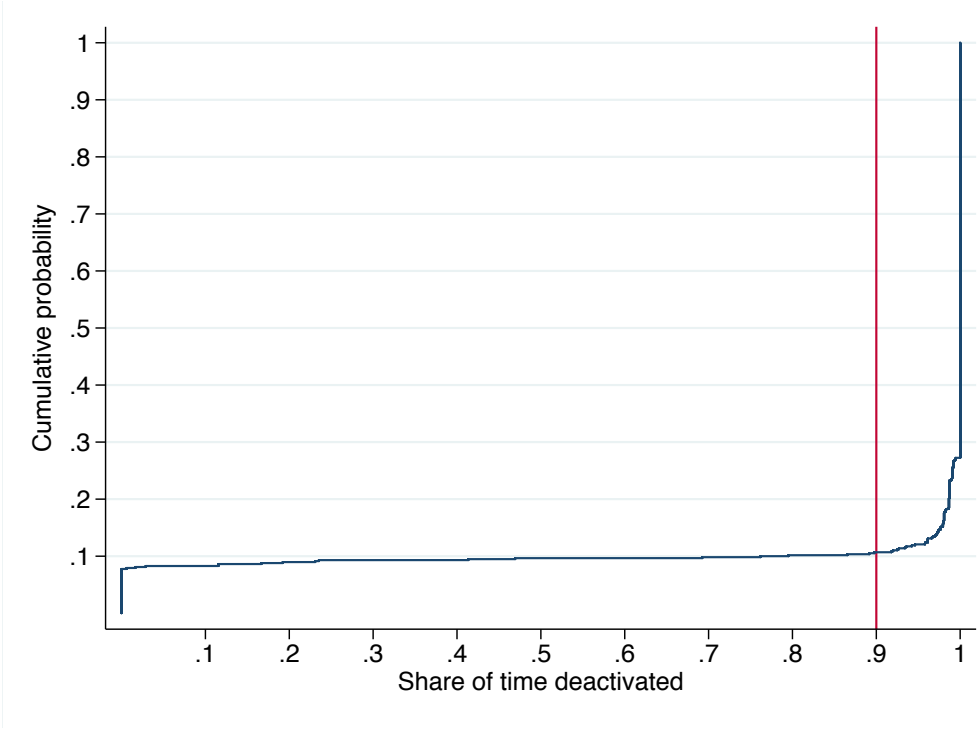
Opinions about Facebook	
Improves social life	To what extent do you think Facebook improves or worsens people's social lives?
Good for you	To what extent do you think Facebook is good or bad for you?
Good for society	To what extent do you think Facebook is good or bad for society?
Makes people happy	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."?
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?
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]
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)
Post-endline mobile use	[if have an iPhone] Please write down the amount of screen time you spent on the Facebook app according to your battery report. [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?
Secondary outcomes	

Voted Republican	If the elections for U.S. 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] [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: "Active" users often post status updates, comment on other people's walls and pictures, post photos, etc. "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. Which would you say describes your Facebook use best? 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?

D Additional Empirical Results

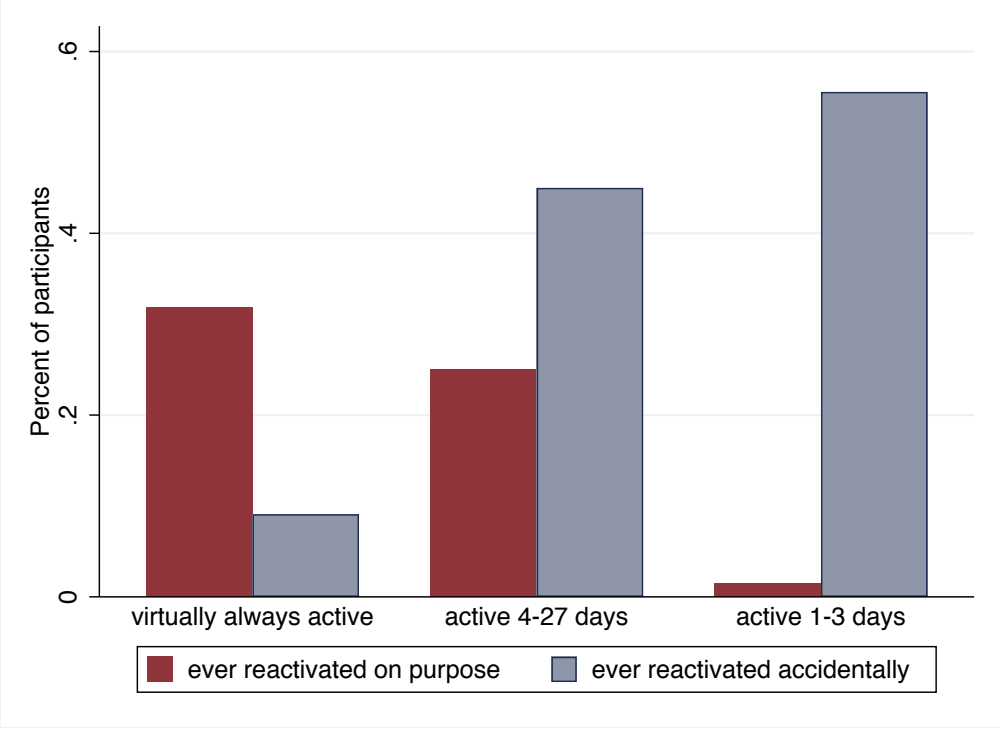
D.A Compliance with Deactivation

Figure A5: **Treatment Group Distribution of Share of Time Deactivated**



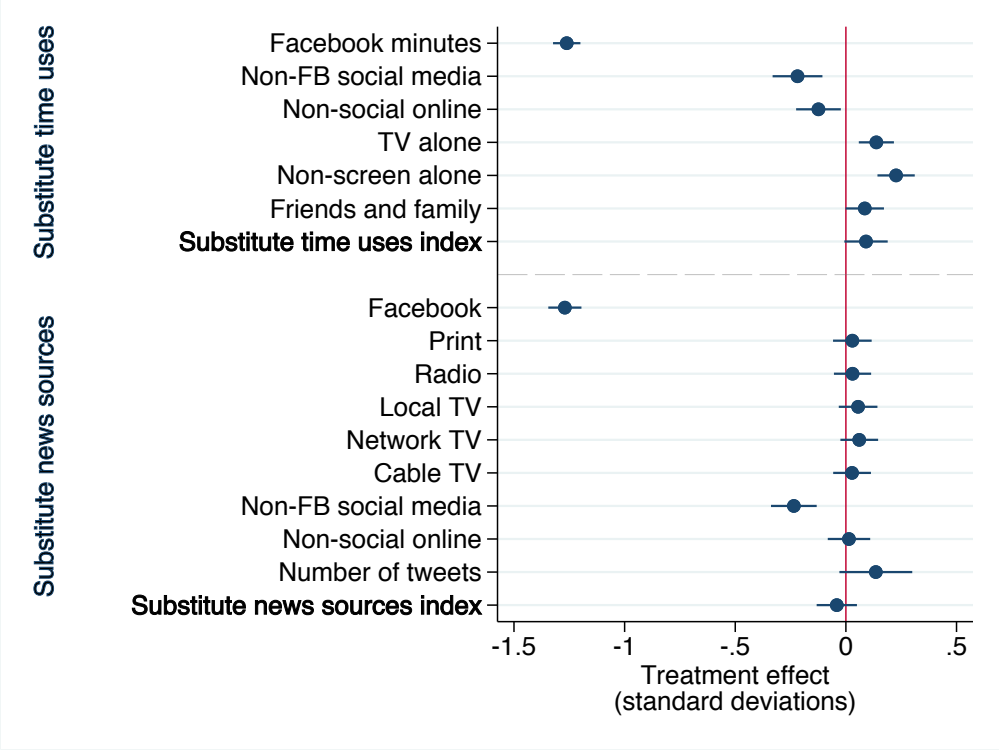
Notes: For each individual in the Treatment group who was willing to accept less than $102 to deactivate Facebook for the four weeks after midline, we calculate the share of the deactivation checks in which that person was deactivated. This figure presents the cumulative distribution of the share of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

Figure A6: **Reasons for Failure to Deactivate**



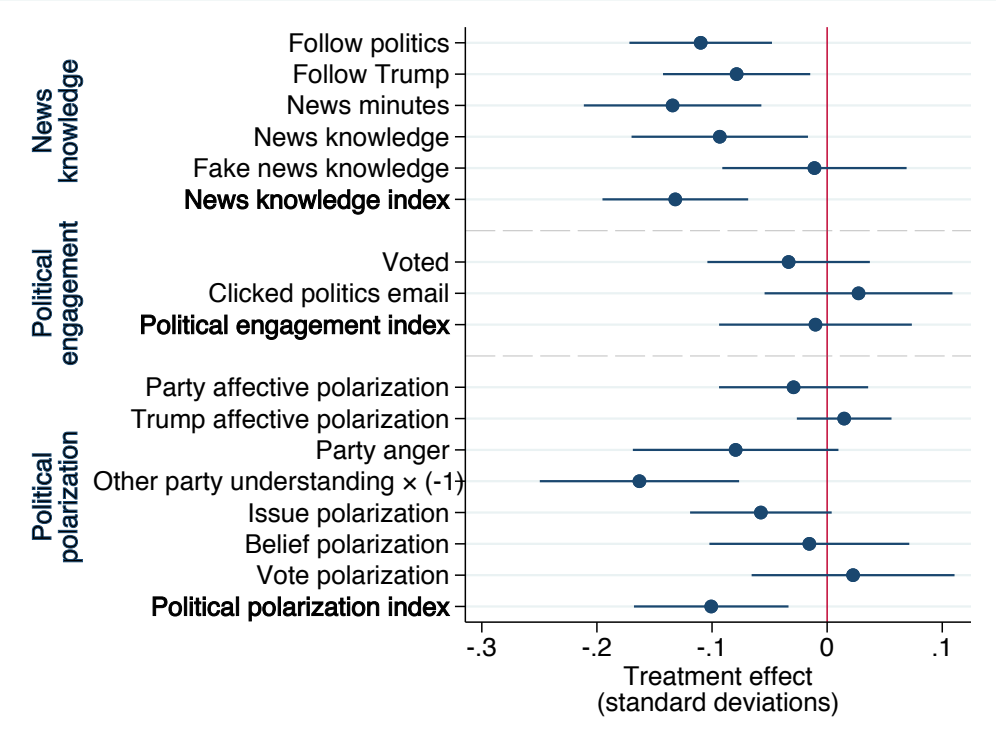
D.B Treatment Effect Estimates Using Equation (2)

Figure A7: **Substitutes for Facebook Using Equation (2)**



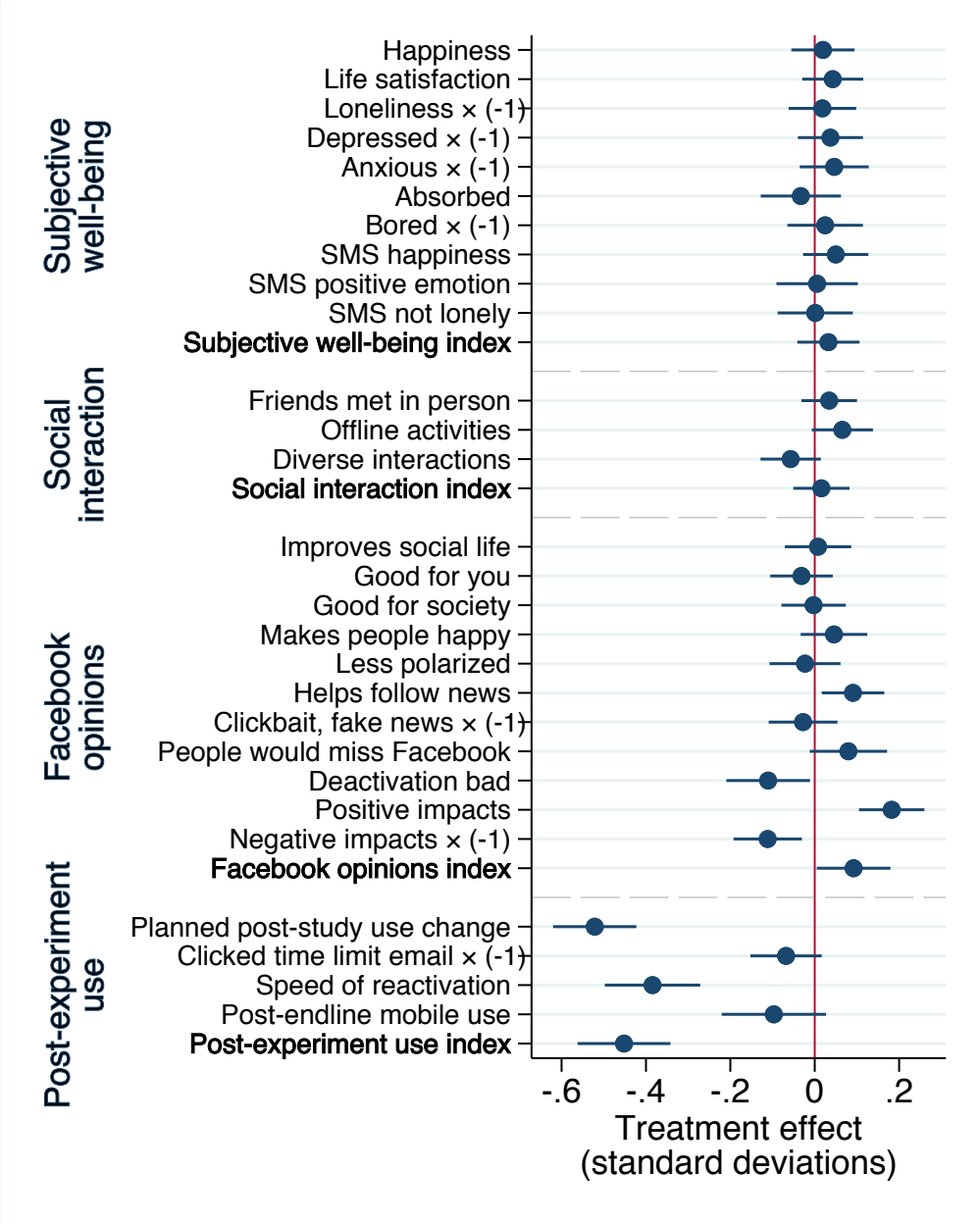
Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (2). All variables are normalized so that the control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section II.C for variable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

Figure A8: **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 II.C for variable definitions.

Figure A9: **Effects on Subjective Well-Being, Social Interactions, and Opinions about Facebook Using Equation (2)**

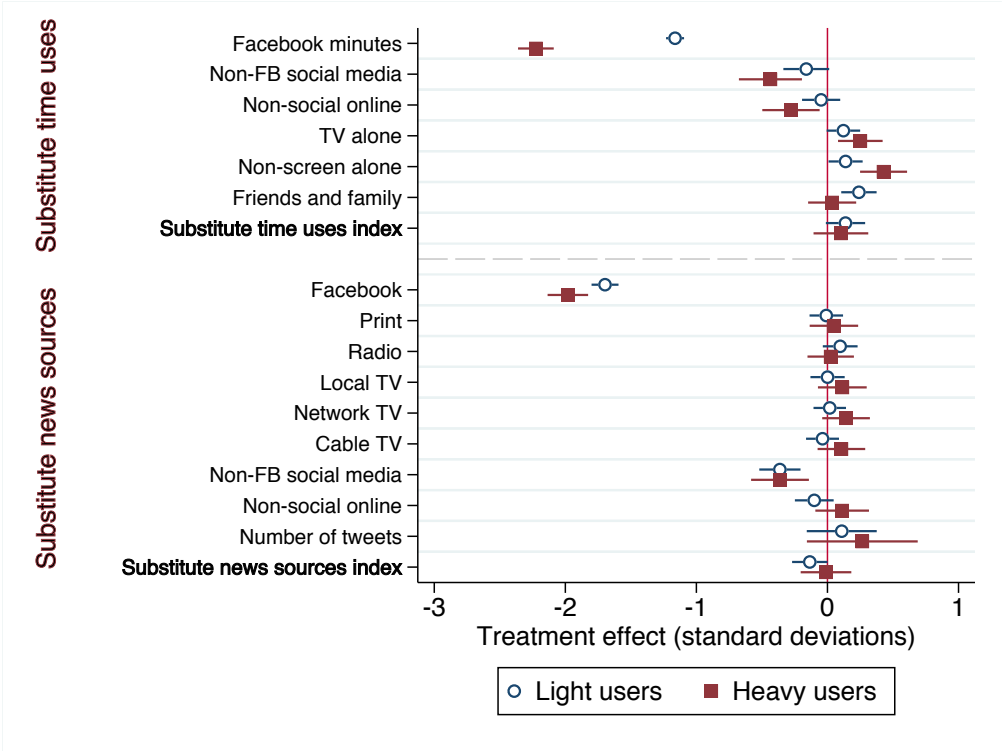


Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (2). All variables are normalized so that the control group endline distribution has a standard deviation of one. Error bars reflect 95 percent confidence intervals. See Section II.C for variable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

D.C Heterogeneous Treatment Effects

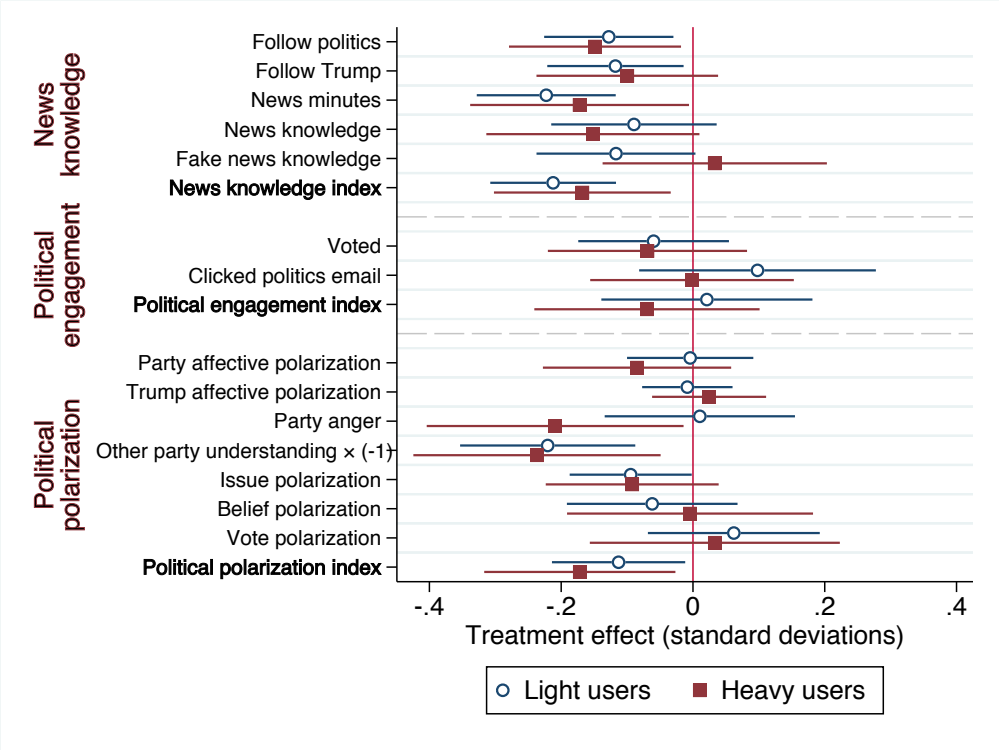
D.C.1 Light and Heavy Users

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



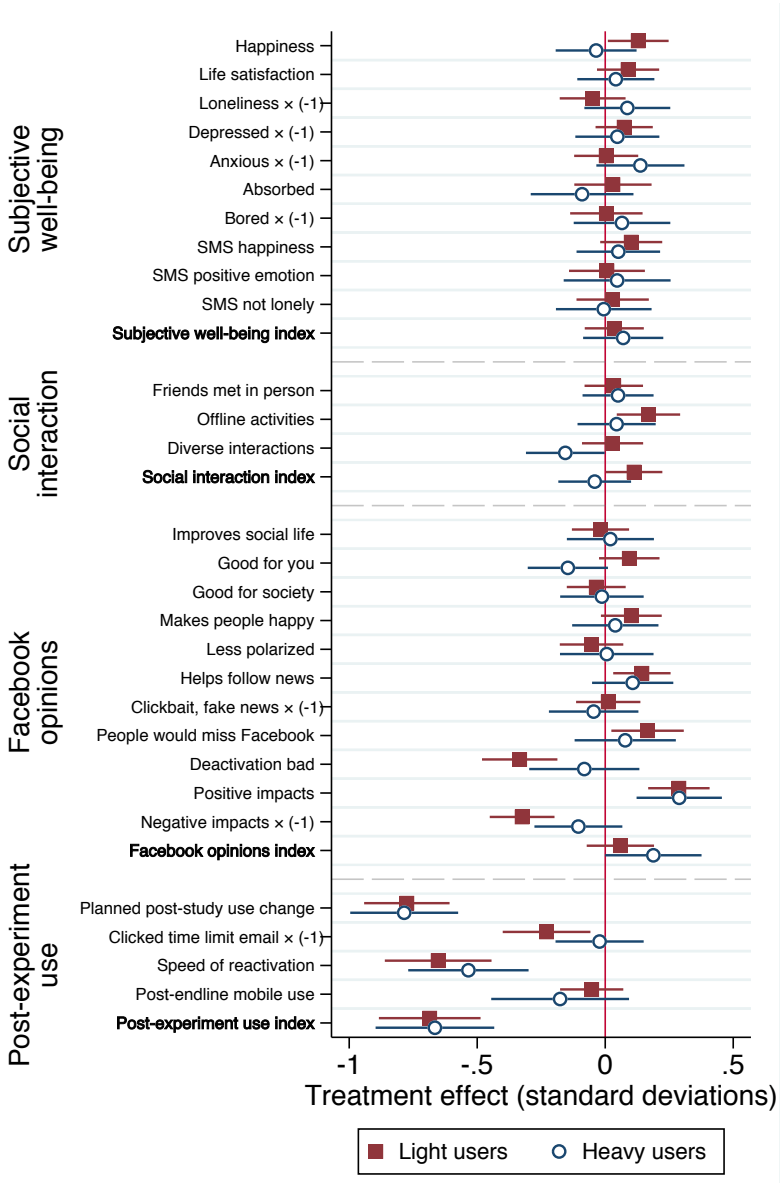
Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 75 daily minutes, the median amount of Facebook use in the impact evaluation sample. All variables are normalized so that the control group endline distribution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

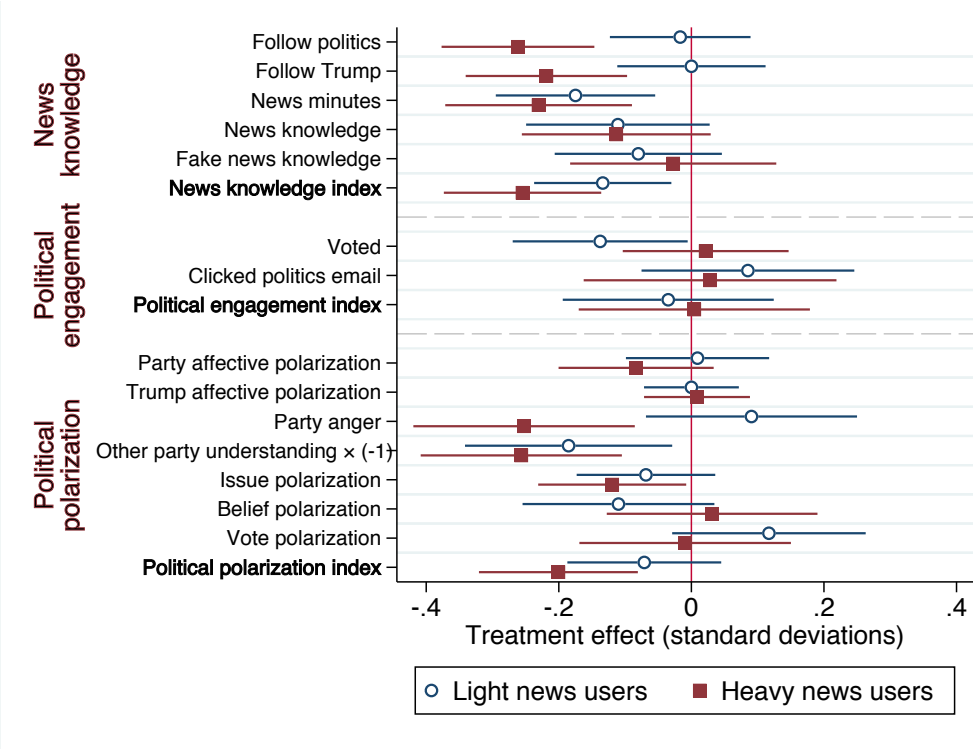
Figure A12: **Effects on Subjective Well-Being, Social Interactions, and Opinions about Facebook for Light and Heavy Users**



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 II.C for variable definitions.

D.C.2 Light and Heavy News Users

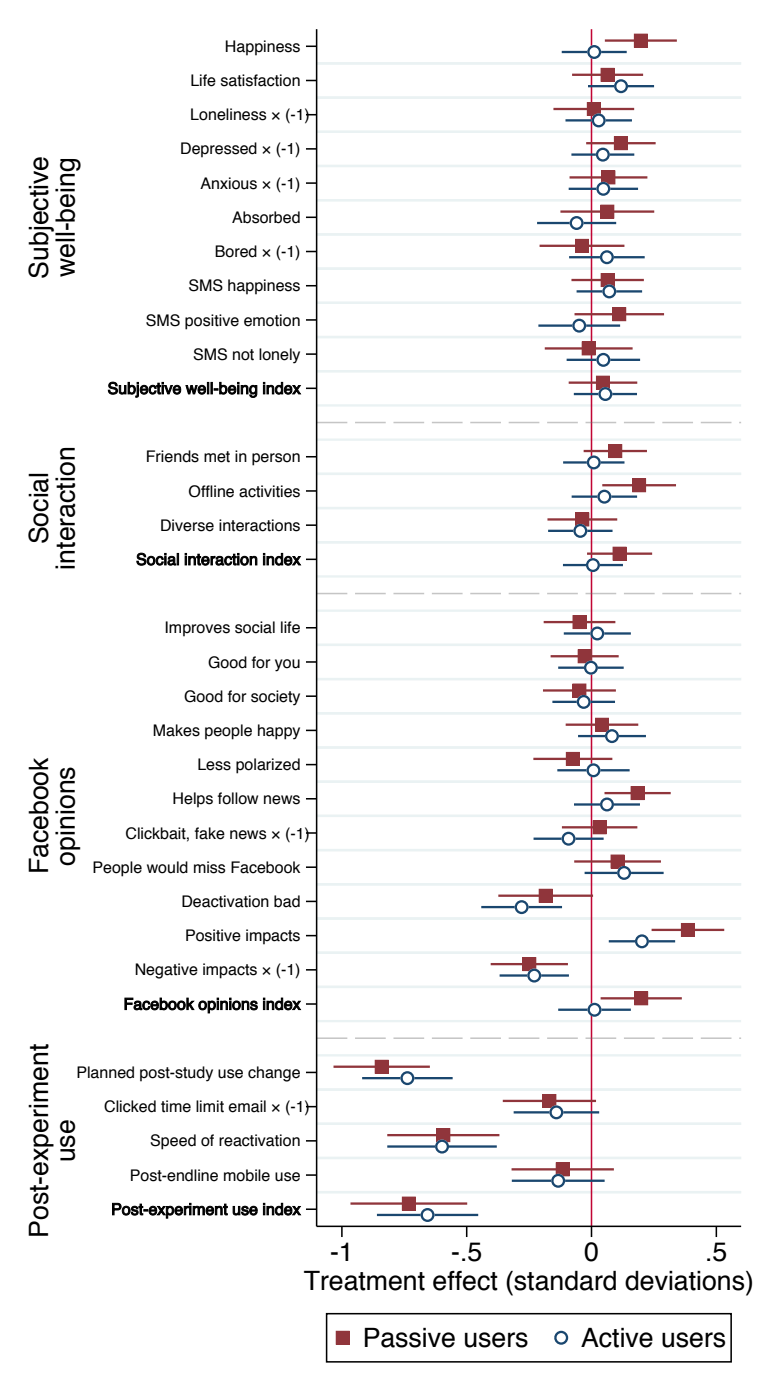
Figure A13: **Effects on News and Political Outcomes for Light and Heavy News Users**



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 II.C for variable definitions.

D.C.3 Active and Passive Users

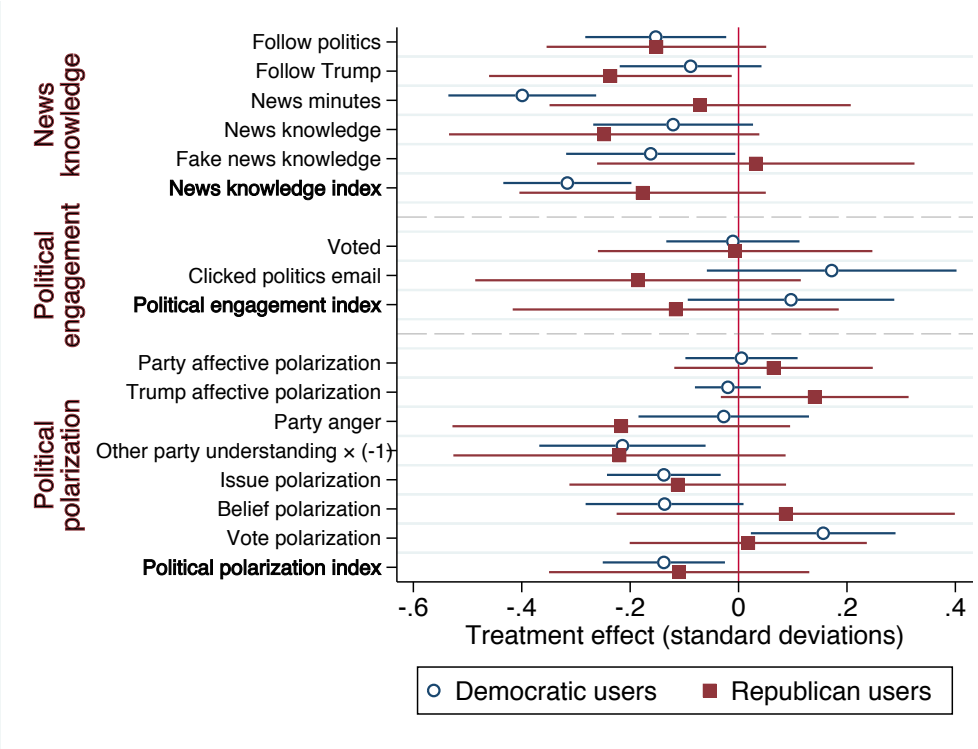
Figure A14: **Effects on Subjective Well-Being, Social Interactions, and Opinions about Facebook for Active and Passive Users**



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 II.C for variable definitions.

D.C.4 Democrats and Republicans

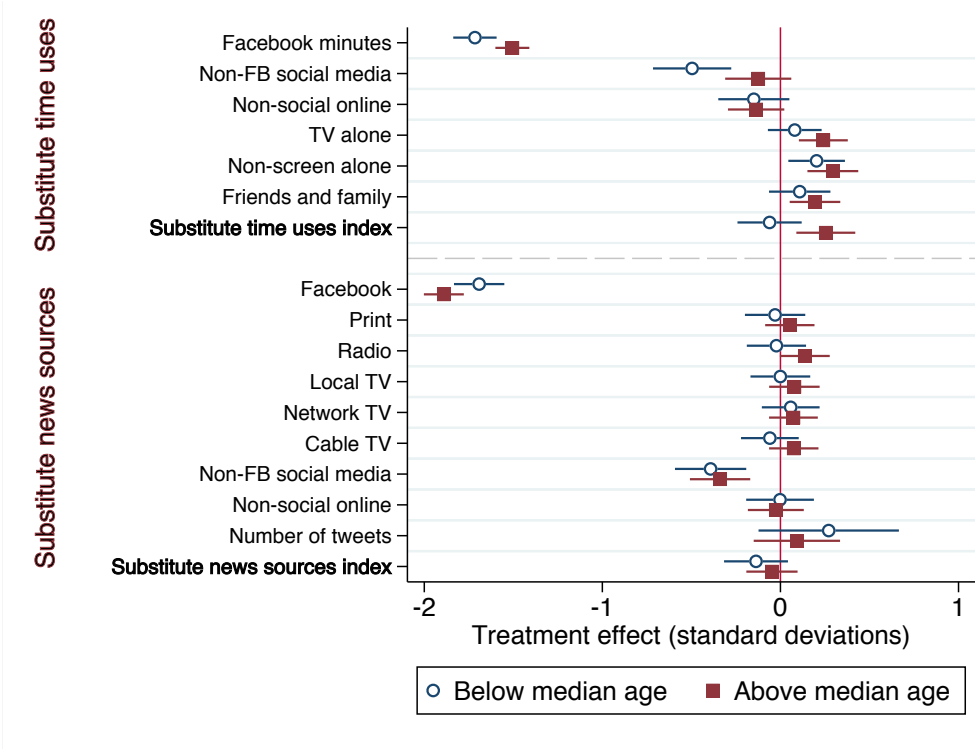
Figure A15: **Effects on News and Political Outcomes for Democrats and Republicans**



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 II.C for variable definitions.

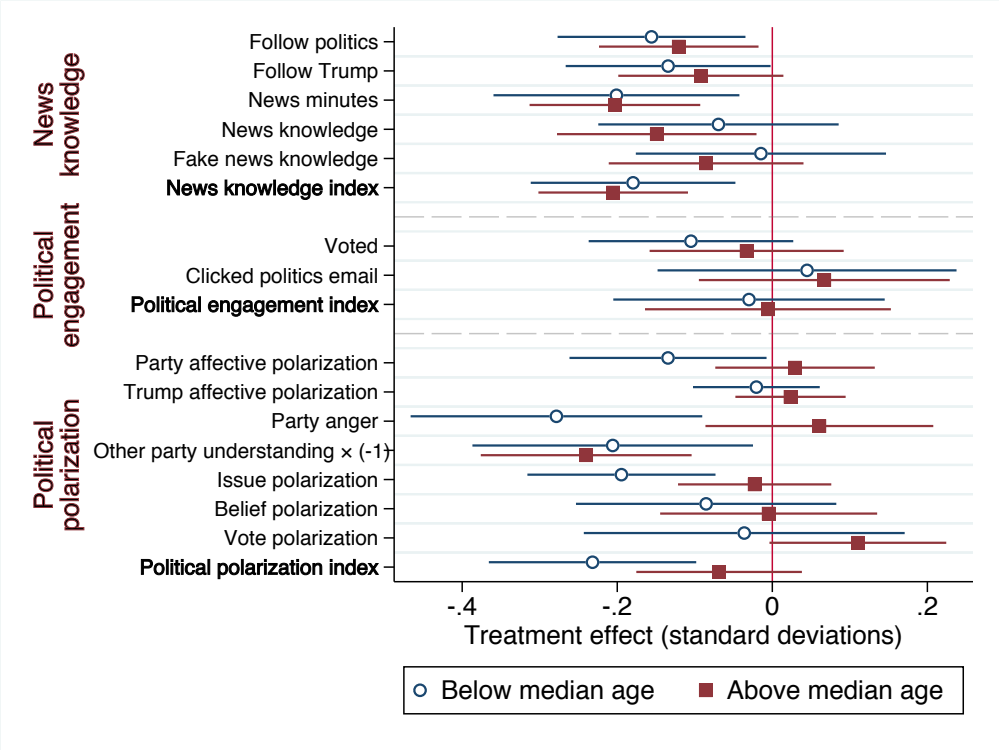
D.C.5 Younger and Older Users

Figure A16: **Substitutes for Facebook for Younger and Older Users**



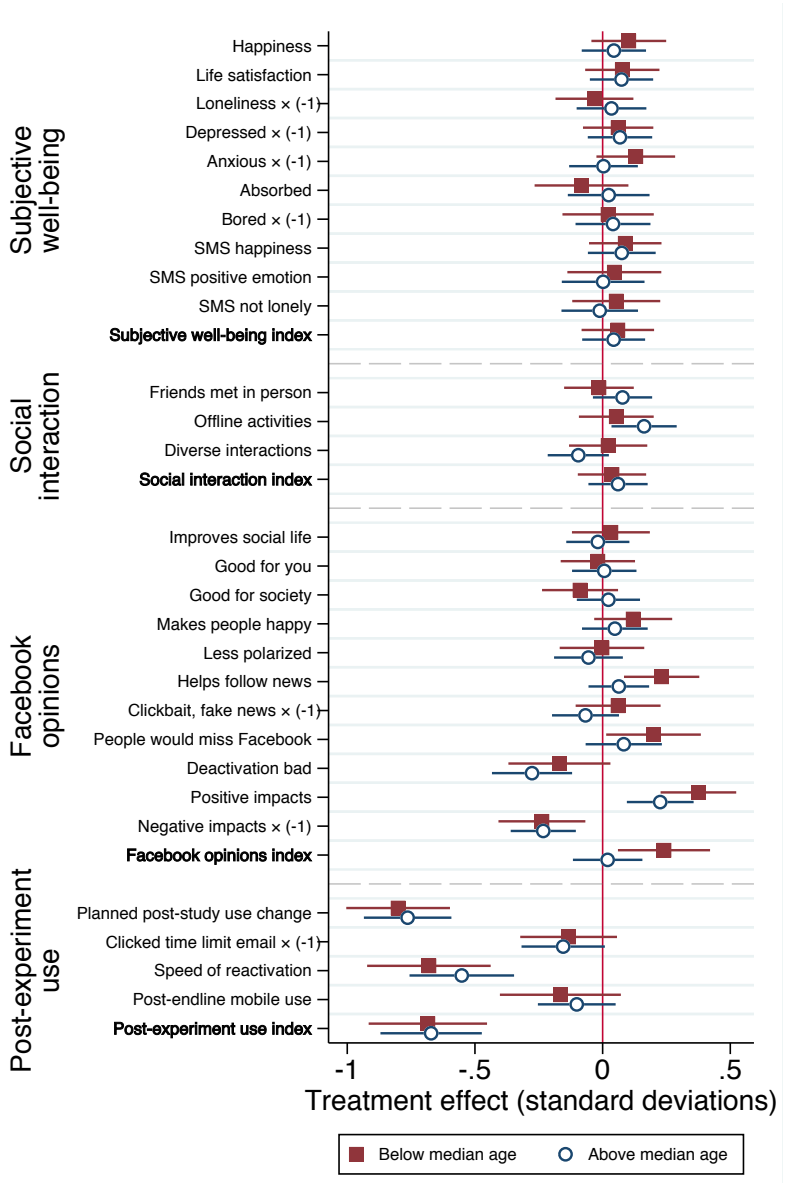
Notes: This figure presents local average treatment effects of Facebook deactivation estimated using Equation (1) for participants above vs. below 31.5 years, the median age in the impact evaluation sample. All variables are normalized so that the control group endline distribution has a standard deviation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

Figure A17: **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 II.C for variable definitions.

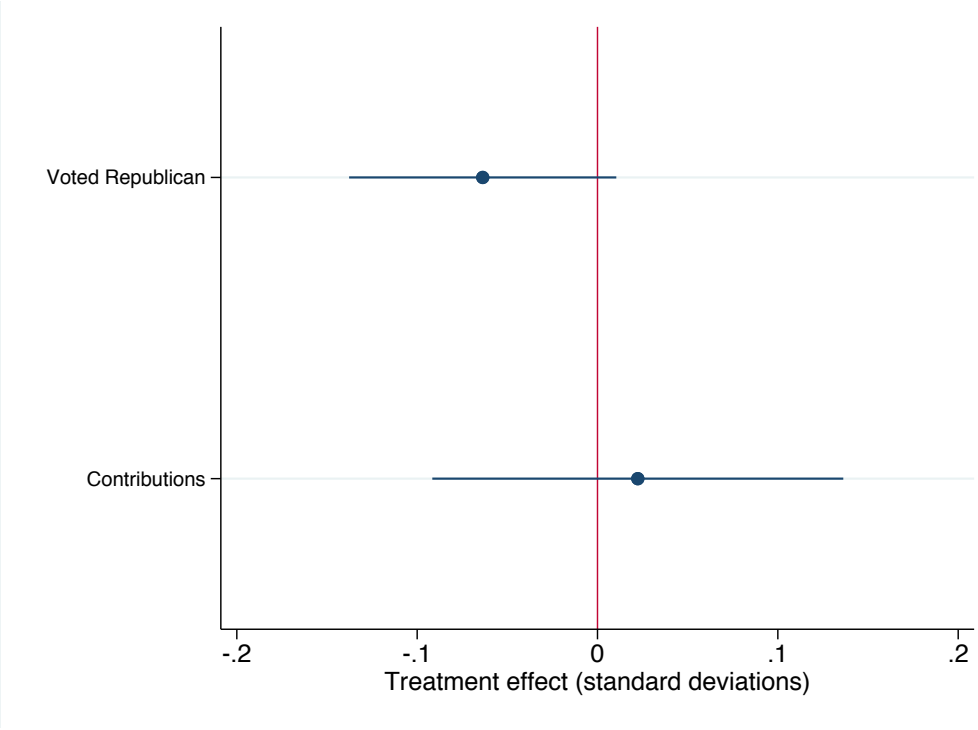
Figure A18: **Effects on Subjective Well-Being, Social Interactions, and Opinions about Facebook for Younger and Older Users**



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 II.C for variable definitions.

D.D Treatment Effects on Secondary Outcomes

Figure A19: **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 II.C for variable definitions.

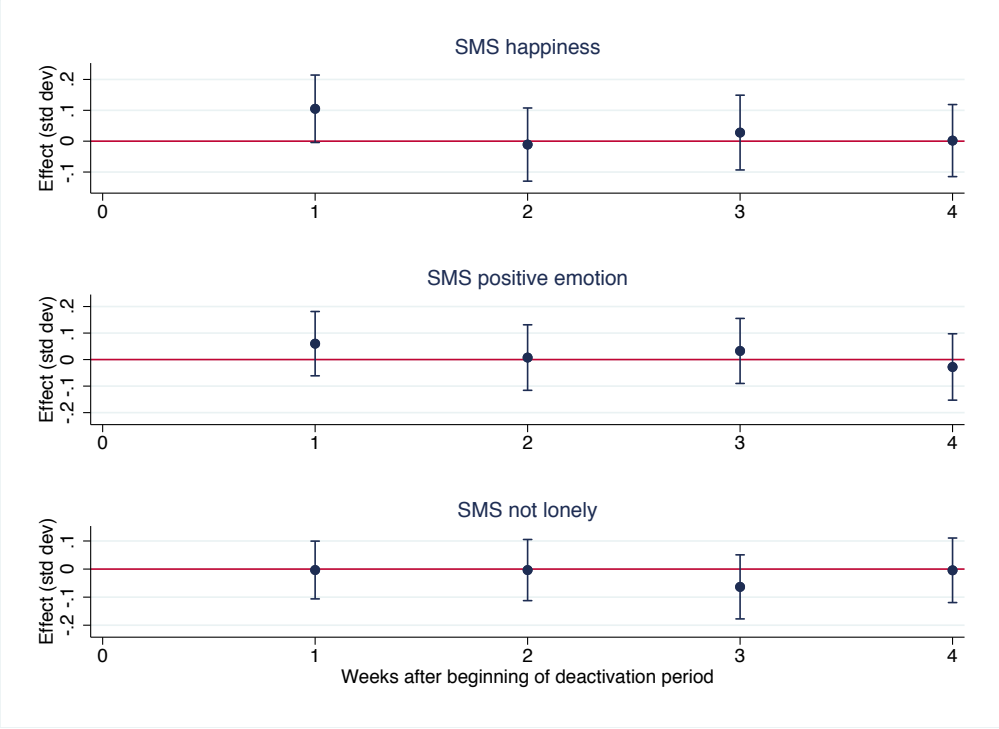
D.E Additional Empirical Results

Table A3: **Effects of Facebook Use on Selected Un-Normalized Outcomes**

	(1) Endline daily Facebook minutes	(2) Post-endline daily Facebook minutes in mobile app
Equation (1)	-59.63	-9.88
Equation (2)	-47.22	-8.50

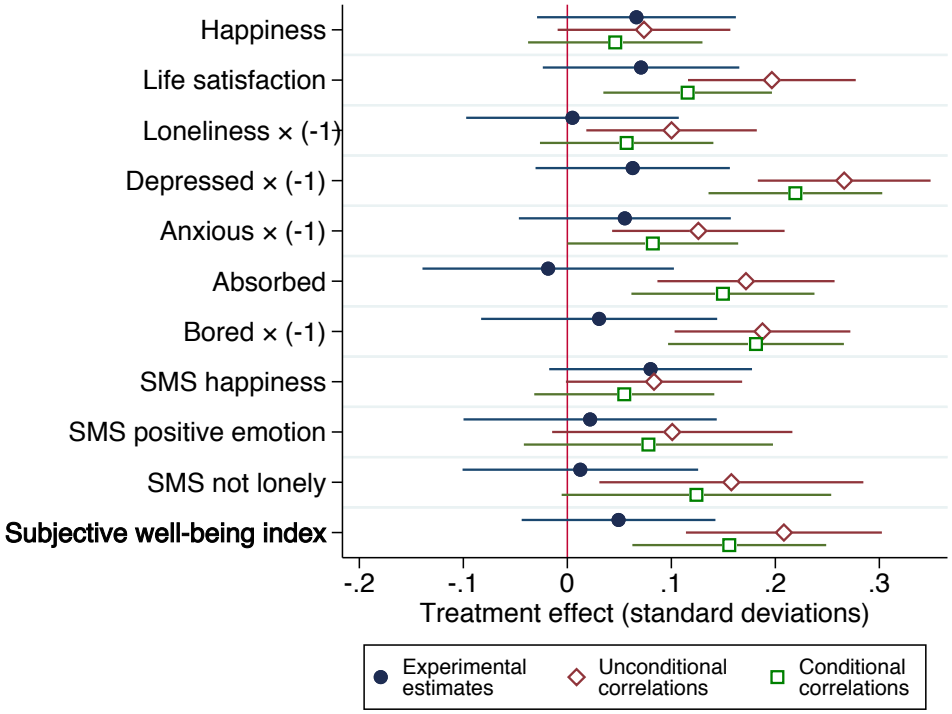
Notes: This table presents local average treatment effects of Facebook deactivation on daily average minutes of Facebook use estimated using Equation (1) and Equation (2). Column 1 presents estimates of effects on Facebook use reported at endline. Column 2 presents estimates of effects on Facebook use in the mobile app only, reported in the post-endline survey. Column 3 presents estimates of effects on daily average minutes spent consuming news.

Figure A20: **Effects on Subjective Well-Being Measured in Text Messages**



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 II.C for variable definitions.

Figure A21: **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 600 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 II.C for variable definitions.