

## The Welfare Effects of Social Media<sup>†</sup>

By HUNT ALLCOTT, LUCA BRAGHERI, SARAH EICHMEYER,  
AND MATTHEW GENTZKOW\*

*The rise of social media has provoked both optimism about potential societal benefits and concern about harms such as addiction, depression, and political polarization. In a randomized experiment, we find that deactivating Facebook for the four weeks before the 2018 US midterm election (i) reduced online activity, while increasing offline activities such as watching TV alone and socializing with family and friends; (ii) reduced both factual news knowledge and political polarization; (iii) increased subjective well-being; and (iv) caused a large persistent reduction in post-experiment Facebook use. Deactivation reduced post-experiment valuations of Facebook, suggesting that traditional metrics may overstate consumer surplus. (JEL D12, D72, D90, I31, L82, L86, Z13)*

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

\*Allcott: New York University and NBER (email: [hunt.allcott@nyu.edu](mailto:hunt.allcott@nyu.edu)); Braghieri: Stanford University (email: [lucabrag@stanford.edu](mailto:lucabrag@stanford.edu)); Eichmeyer: Stanford University (email: [saraeich@stanford.edu](mailto:saraeich@stanford.edu)); Gentzkow: Stanford University and NBER (email: [gentzkow@stanford.edu](mailto:gentzkow@stanford.edu)). Esther Duflo was the coeditor for this article. We thank Nancy Baym, Moira Burke, Annie Franco, Alex Leavitt, Todd Rogers, Joel Waldfogel, and seminar participants at Berkeley, Centro de Investigación y Docencia Económicas, Chicago, Columbia, Instituto Tecnológico Autónomo de México, Microsoft, NBER Digitization, NYU, Stanford, the Technology Policy Institute, UCLA, UC Santa Cruz, and University of Washington for helpful comments. We thank Raj Bhargava, Zong Huang, and Kelly B. Wagman for exceptional research assistance. We are grateful to the Sloan Foundation and the Knight Foundation for generous support. The study was approved by Institutional Review Boards at Stanford (eProtocol 45403) and NYU (IRB-FY2018-2139). This RCT was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0003409. 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 perform 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.

<sup>†</sup>Go to <https://doi.org/10.1257/aer.20190658> to visit the article page for additional materials and author disclosure statements.

well-being (Myers 2000; Reis, Collins, and Berscheid 2000; Argyle 2001; Chopik 2017), this could be expected to bring widespread improvements to individual welfare. Many have also pointed to wider social benefits, from facilitating protest and resistance in autocratic countries, to encouraging activism and political participation in established democracies (Howard et al. 2011, Kirkpatrick 2011).

More recent discussion has focused on an array of possible negative impacts. At the individual level, many have pointed to negative correlations between intensive social media use and both subjective well-being and mental health.<sup>1</sup> 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.<sup>2</sup> Alter (2018) and Newport (2019), along with other academics and prominent Silicon Valley executives in the “time well-spent” movement, argue that digital media devices and social media apps are harmful and addictive. At the broader social level, concern has focused particularly on a range of negative political externalities. Social media may create ideological “echo chambers” among like-minded friend groups, thereby increasing political polarization (Sunstein 2001, 2017; Settle 2018). Furthermore, social media are the primary channel through which misinformation spreads online (Allcott and Gentzkow 2017), and there is concern that coordinated disinformation campaigns can affect elections in the United States and abroad.

In this paper, we report on a large-scale randomized evaluation of the welfare impacts of Facebook, focusing on US users in the run-up to the November 2018 midterm elections. We recruited a sample of 2,743 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 61 percent of these subjects with WTA less than \$102 to either a Treatment group that was paid to deactivate, or a Control group that was not. We verified compliance with deactivation by regularly checking participants’ public profile pages. We measured a suite of outcomes using text messages, surveys, emails, direct measurement of Facebook and Twitter activity, and administrative voting records. Less than 2 percent of the sample failed to complete the endline survey, and the Treatment group’s compliance with deactivation exceeded 90 percent.

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

<sup>1</sup> See, for example, Vanden Abeele et al. (2018); Burke and Kraut (2016); Ellison, Steinfield, and Lampe (2007); Frison and Eggermont (2015); Kross et al. (2013); Satıcı and Uysal (2015); Shakyia and Christakis (2017); and Tandoc, Ferrucci, and Duffy (2015). See Appel, Gerlach, and Crusius (2016) and Baker and Algorta (2016) for reviews.

<sup>2</sup> See, for example, Twenge, Sherman, and Lyubomirsky (2016); Twenge and Park (2019); Twenge, Martin, and Campbell (2018); and Twenge et al. (2018).

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

Our second set of results focuses on political externalities, proxied by news knowledge, political engagement, and political polarization. Consistent with the reported reduction in news consumption, we find that Facebook deactivation significantly reduced news knowledge and attention to politics. The Treatment group was less likely to say they follow news about politics or the President, and less able to correctly answer factual questions about recent news events. Our overall index of news knowledge fell by 0.19 standard deviations. There is no detectable effect on political engagement, as measured by voter turnout in the midterm election and the likelihood of clicking on email links to support political causes. Deactivation significantly reduced polarization of views on policy issues and a measure of exposure to polarizing news. Deactivation did not statistically significantly reduce affective polarization (i.e., negative feelings about the other political party) or polarization in factual beliefs about current events, although the coefficient estimates also point in that direction. Our overall index of political polarization fell by 0.16 standard deviations. As a point of comparison, prior work has found that a different index of political polarization rose by 0.38 standard deviations between 1996 and 2018 (Boxell 2018).

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

<sup>3</sup>Correlation studies on active versus passive Facebook use include Burke, Marlow, and Lento (2010); Burke, Kraut, and Marlow (2011); Burke and Kraut (2014); and Krasnova et al. (2013), and randomized experiments include Deters and Mehl (2013) and Verduyn et al. (2015).

Our fourth set of results considers whether deactivation affected people's demand for Facebook after the study was over, as well as their opinions about Facebook's role in society. As the experiment ended, participants reported planning to use Facebook much less in the future. Several weeks later, the Treatment group's reported usage of the Facebook mobile app was about 11 minutes (22 percent) lower than in Control. The Treatment group was more likely to click on a post-experiment email providing information about tools to limit social media usage, and 5 percent of the Treatment group still had their accounts deactivated nine weeks after the experiment ended. Our overall index of post-experiment Facebook use is 0.61 standard deviations lower in Treatment than in Control. In response to open-answer questions several weeks after the experiment ended, the Treatment group was more likely to report that they were using Facebook less, had uninstalled the Facebook app from their phones, and were using the platform more judiciously. Reduced post-experiment use aligns with our finding that deactivation improved subjective well-being, and it is also consistent with the hypotheses that Facebook is habit forming in the sense of Becker and Murphy (1988) or that people learned that they enjoy life without Facebook more than they had anticipated.

Deactivation caused people to appreciate Facebook's both positive and negative impacts on their lives. Consistent with our results on news knowledge, the Treatment group was more likely to agree that Facebook helps people to follow the news. About 80 percent of the Treatment group agreed that deactivation was good for them, but they were also more likely to think that people would miss Facebook if they used it less. In free response questions, the Treatment group wrote more text about how Facebook has both positive and negative impacts on their lives. The opposing effects on these specific metrics cancel out, so our overall index of opinions about Facebook is unaffected.

Our work also speaks to an adjacent set of questions around how to measure the economic gains from free online services such as search and media.<sup>4</sup> In standard models with consumers who correctly optimize their allocation of time and money, researchers can approximate the consumer surplus from these services by measuring time use or monetary valuations, as in Brynjolfsson and Oh (2012); Brynjolfsson, Eggers, and Gannamaneni (2018); Corrigan et al. (2018); and others. But if users do not understand the ways in which social media could be addictive or make them unhappy, these standard approaches could overstate consumer surplus gains. Sagioglu and Greitemeyer (2014) provides suggestive evidence: while their participants predicted that spending 20 minutes on Facebook would make them feel better, it actually caused them to feel worse. Organizations such as Time to Log Off argue that a 30-day "digital detox" would help people align their social media usage with their own best interest.

To quantify the possibility that deactivation might help the Treatment group to understand ways in which their use had made them unhappy, we elicited willingness-to-accept at three separate points, using incentive-compatible Becker-DeGroot-Marschak (1964) mechanisms. First, on October 11, we elicited WTA to deactivate Facebook for weeks 1–4 of the experiment, between October 12

<sup>4</sup> See, for example, Brynjolfsson and Saunders (2009); Byrne, Fernald, and Reinsdorf (2016); Nakamura, Samuels, and Soloveichik (2016); Brynjolfsson, Rock, and Syverson (2019); and Syverson (2017).

and November 8. We immediately told participants the amount that they had been offered to deactivate (\$102 for the Treatment group, \$0 for Control), and thus whether they were expected to deactivate over that period. We then immediately elicited WTA to deactivate Facebook for the next four weeks *after* November 8, i.e., weeks 5–8. When November 8 arrived, we then re-elicited WTA to deactivate for weeks 5–8. The Treatment group's change in valuation for weeks 5–8 reflects a time effect plus the effect of deactivating Facebook. The Control group's parallel valuation change reflects only a time effect. Thus, the difference between how Treatment versus Control change their WTAs for deactivation for weeks 5–8 reflects projection bias, learning, or other unanticipated experience effects from deactivation.<sup>5</sup>

After weighting our sample to match the average US Facebook user on observables, the median and mean willingness-to-accept to deactivate Facebook for weeks 1–4 were \$100 and \$180, respectively. These valuations are larger than most estimates in related work by Brynjolfsson, Eggers, and Gannamaneni (2018); Corrigan et al. (2018); Mosquera et al. (2018); and Sunstein (forthcoming). A standard consumer surplus calculation would aggregate the mean valuation across the estimated 172 million US Facebook users, giving \$31 billion in consumer surplus from four weeks of Facebook. However, consistent with our other results that deactivation reduced demand for Facebook, deactivation caused WTA for weeks 5–8 to drop by up to 14 percent. This suggests that traditional consumer surplus metrics overstate the true welfare gains from social media, though a calculation that adjusts for the downward WTA revision would still imply that Facebook generates enormous flows of consumer surplus.

What do our results imply about the overall net welfare impact of Facebook? On the one hand, Facebook deactivation increased subjective well-being, and 80 percent of the Treatment group reported that deactivation was good for them. On the other hand, participants were unwilling to give up Facebook unless offered fairly large amounts of money: even after they had deactivated for four weeks, which should have allowed at least some learning or “detox” from addiction. It is not entirely clear whether one should prioritize the survey measures or monetary valuations as normative measures of consumer welfare. Benjamin et al. (2012) suggests that subjective well-being measures like ours are not a complete measure of what people are trying to maximize when they make decisions, but Bohm, Lindén, and Sonnégård (1997); Mazar, Köszegi, and Ariely (2014); and other studies make clear that monetary valuations are not closely held and can be easily manipulated. We think of these tensions as fodder for future research.

Our results should be interpreted with caution, for several reasons. First, effects could differ with the duration, time period, or scale of deactivation. A longer period without Facebook might have less impact on news knowledge as people find alternative news sources, and either more or less impact on subjective well-being. Effects might be different for our pre-election deactivation than for deactivation in other periods. Furthermore, the effects of deactivating a large share of Facebook users

<sup>5</sup>This measurement connects to the literature on habit formation and projection bias, including Acland and Levy (2015); Becker and Murphy (1988); Becker, Grossman, and Murphy (1991); Busse et al. (2015); Charness and Gneezy (2009); Conlin, O'Donoghue, and Vogelsang (2007); Fujiwara, Meng, and Vogl (2016); Gruber and Köszegi (2001); Hussam et al. (2016); Loewenstein, O'Donoghue, and Rabin (2003); and Simonsohn (2010).



would likely be different due to network effects, so our parameters are most relevant for individuals independently determining their own Facebook use. Second, our sample is not fully representative. Our participants are relatively young, well-educated, and left-leaning compared to the average Facebook user; we included only people who reported using Facebook more than 15 minutes per day; and people willing to participate in our experiment may also differ in unobservable ways. Third, many of our outcome variables are self-reported, adding scope for both measurement error and experimenter demand effects. However, Section IVF finds no evidence of demand effects, and our non-self-reported outcomes paint a similar picture to the survey responses.

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

Sections I through III present the experimental design, descriptive statistics, and empirical strategy. Section IV presents the impact evaluation, and Section V discusses measurement of the consumer surplus generated by Facebook.

## I. Experimental Design

### A. Experiment Overview

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

Between September 24 and October 3, we recruited participants using Facebook ads. Our ad said, “Participate in an online research study about internet browsing and

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

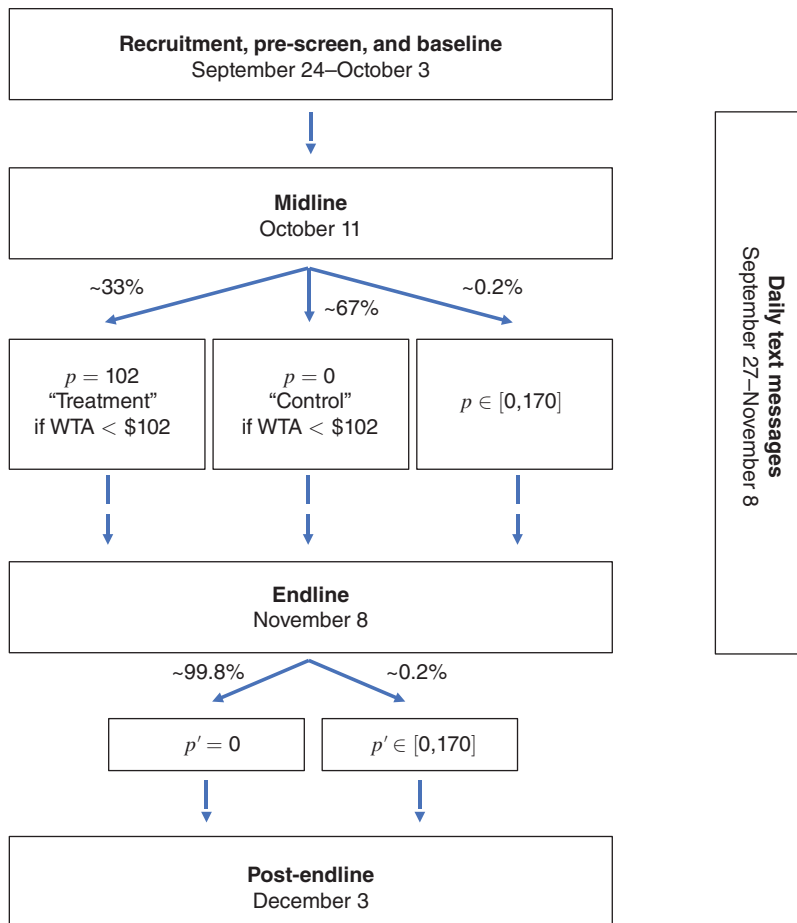


FIGURE 1. EXPERIMENTAL DESIGN

earn an easy \$30 in electronic gift cards.” Online 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. A total of 1,892,191 unique users were shown the ad, of whom 32,201 clicked on it. This 1.7 percent click-through rate is about twice the average click-through rate on Facebook ads across all industries.<sup>7</sup>

Clicking on the ad took the participant to a brief pre-screen survey, which included several background demographic questions and the consent form. A total of 17,335 people passed the pre-screen, by reporting being a US resident born between the years 1900 and 2000 who uses Facebook more than 15 minutes and no more than 600 minutes per day. Of those people, 7,455 consented to participate in the study.

After completing the consent form, participants began the baseline survey. The baseline recorded email addresses, additional demographics, and a range of outcome

<sup>7</sup> Mark Irvine, “Facebook Ad Benchmarks for YOUR Industry,” *WordStream*, August 27, 2019, <https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks>.

variables. We also asked for each participant's name, zip code, Twitter handle, and phone number ("in order for us to send you text messages during the study"), as well as the URL of their Facebook profile page (which we would use "solely to observe whether your Facebook account is active"). Finally, we informed people that we would later ask them to deactivate their accounts for two 24-hour periods, and confirmed their willingness to do so. (We required all participants regardless of treatment status to deactivate for these 24-hour periods to minimize selective attrition and to ensure that the valuations described below reflect value of Facebook access, not the fixed cost of the deactivation process.)

In all, 3,910 people finished the baseline survey and were willing to deactivate. Of those, 1,013 were dropped from the experiment 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 versus baseline survey, completing the survey in less than ten minutes, no text in short-answer boxes, and other patterns suggesting careless responses). The remaining 2,897 participants had valid baseline data, were included in our stratified randomization, and were invited to take the midline survey.

On October 11, we sent an email invitation to the midline survey. The survey first asked participants to deactivate their Facebook accounts for 24 hours and guided them through the process. The survey clearly explained what deactivation entailed and how we would monitor deactivation. Facebook allows users to deactivate and reactivate their accounts at any time. We informed participants that they could continue to use Facebook Messenger while deactivated, and that their profile and friend network would be unchanged when they reactivated. We emphasized that Facebook would automatically reactivate their account if they logged into the Facebook website or app, or if they actively logged into any *other* app using their Facebook login credentials.<sup>8</sup> We informed participants that "We will verify whether or not you deactivated your account by pinging the Facebook URL" that they had provided in the baseline survey.

The midline survey then used a Becker-DeGroot-Marschak (BDM) mechanism to elicit willingness-to-accept (WTA) to stay deactivated for four weeks rather than 24 hours.<sup>9</sup> We then revealed the BDM price offer. An additional 154 participants had dropped out before this point of the midline survey, leaving 2,743 who received their price offer. Participants whose WTA was strictly less than the price draw were informed that they should deactivate for the full four weeks after midline. Finally, the midline survey reminded people that we would again ask them to deactivate for

<sup>8</sup> A user's Facebook account automatically reactivates whenever the user actively logs into any other app using their Facebook login credentials. However, this does not fully preclude people from using other apps for which they had used Facebook to log in. People can continue using other apps if they are already logged in, can set up non-Facebook logins, or can log in with Facebook and then again deactivate their Facebook account.

<sup>9</sup> The survey explained, "The computer has randomly generated an amount of money to offer you to deactivate your Facebook account for the next 4 weeks. Before we tell you what the offer is, we will ask you the smallest offer you would be willing to accept. If the offer the computer generated is above the amount you give, we will ask you to deactivate for 4 weeks and pay you the offered amount if you do. If the offer is below that amount, we will not ask you to deactivate." We then asked several comprehension questions to make sure that participants understood the mechanism. We did not tell participants the distribution or support of the offer prices, both because we did not want to artificially truncate the distribution of elicited WTA and because prior studies have found that providing information on the bounds of the offer price distribution can affect BDM valuations (Bohm, Lindén, and Sonnegård 1997; Mazar, Kőszegi, and Ariely 2014).



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 refer to the four weeks after midline as “weeks 1–4,” and the four weeks after endline as “weeks 5–8.”

On November 8, two days after the midterm election, we sent an email invitation to the endline survey. The endline survey first measured the same outcome variables as the baseline survey. All questions were identical, with the exception of cases discussed in Section IC, such as using updated news knowledge questions and rephrasing questions about the midterm election to be in the past tense. We then asked all participants to again deactivate their Facebook accounts for the next 24 hours, and again elicited WTA to stay deactivated for the next four weeks (i.e., weeks 5–8) instead of the next 24 hours. Participants were told, “With a 50 percent chance we will require you to abide by the decision you made 4 weeks ago; with 50 percent 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 20, we sent an email with links to information on ways to limit smartphone social media use, and on November 25, we sent an email with links to donate to, volunteer for, or sign petitions related to political causes. Clicks on these emails provide additional non-self-reported measures of interest in reducing social media use and political engagement. Online Appendix Figures A2 and A3 present the two emails.

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

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

We verified deactivation by checking each participant’s Facebook profile page URL regularly at random times. While a user can limit how much content other people can see in their profiles, they cannot hide their public profile page, and the public profile URL returns a valid response if and only if their account is active.<sup>10</sup> This is thus our measure of deactivation. For all participants, we verified deactivation approximately once per day for the seven days before midline and all days between endline and the end of January 2019. Between midline and endline, we verified deactivation approximately four times per day for people who were supposed to be

<sup>10</sup>By default, Facebook profile URLs end in a unique number, which is the numeric ID for that person in the Facebook system. Users can update their default URL to be something customized, and they can change their customized URL as often as they want. In the baseline survey, participants reported their profile URLs, which could have been either the default or customized version. Shortly after the baseline survey, we checked if each participant’s Facebook profile URL was valid by pinging it and looking in the page source for the string containing the person’s numeric ID. If the numeric ID existed, we knew that the URL was valid. After that point, we used participants’ numeric IDs to construct their default numeric URLs, which allowed us to correctly measure deactivation even if they changed their customized URL.

deactivated (i.e., the Treatment group) and once every four days for everyone else. 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.<sup>11</sup> These payments were in addition to the \$102 that the Treatment group received in exchange for deactivation.

### B. Randomization

We used the BDM mechanism described above to randomly assign participants to Facebook deactivation. Figure 1 illustrates the randomization. Participants with valid baseline data were randomized into three groups that determined the BDM offer price  $p$  for deactivation in weeks 1–4 (i.e., the weeks between midline and endline):  $p = \$102$  (approximately 33 percent of the sample),  $p = \$0$  (approximately 67 percent), and  $p$  drawn from a uniform distribution on  $[\$0, \$170]$  (approximately 0.2 percent).<sup>12</sup> We balanced the  $p = \$102$  and  $p = \$0$  group assignments within 48 strata defined by age, average daily Facebook use, heavy versus light news use (those who get news from Facebook fairly often or very often versus never, hardly ever, or sometimes), active versus passive Facebook use, and Democrat, Republican, or independent party affiliation.

The effects of Facebook deactivation in weeks 1–4 are identified in the sample of participants who were allocated to  $p = \$102$  or  $p = \$0$  and were willing to accept less than \$102 to deactivate in weeks 1–4. We call this the “impact evaluation sample.” Within the impact evaluation sample, we call  $p = \$102$  the “Treatment” group, and  $p = \$0$  the “Control” group.

For deactivation in weeks 5–8 (i.e., the four weeks after endline), 0.2 percent of participants were randomly selected to a BDM offer price drawn randomly from  $p' \in [0, 170]$ , while the remaining 99.8 percent received offer  $p' = 0$ . We balanced

<sup>11</sup> As discussed above, we did not inform participants of the BDM offer price distribution. Thus, more precisely, truthfully reporting valuations is a strictly dominant strategy only within the support of the offer price distribution that participants expected us to use.

<sup>12</sup> We chose \$102 because our pilot data correctly suggested that there would be a point mass of WTAs at \$100 and that it would maximize statistical power per dollar of cost to set an offer price just high enough to induce those participants to deactivate. We chose \$170 as the top of the uniform distribution because it was the maximum that we could pay participants without requiring tax-related paperwork.

this weeks 5–8 offer price  $p'$  between the weeks 1–4 offer price groups, so two participants who were offered  $p = \$102$  and four participants who were offered  $p = \$0$  were assigned to positive weeks 5–8 offers  $p' \in [0, 170]$ .

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

### C. Outcome Variables

For the impact evaluation, we consider the outcome variables in the nine families described below. Online Appendix B presents survey question text and descriptive statistics for each outcome variable and moderator, grouped by family. We also construct indices that combine the outcome variables within each family, weighting by the inverse of the covariance between variables at endline, as described in Anderson (2008). In constructing these indices, we orient the variables so that more positive values have the same meaning: for example, more positive means “more polarized” in all cases. Outcomes to be multiplied by  $-1$  are followed by “ $\times (-1)$ ” in online Appendix B.

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

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

*Substitute News Sources.*—At baseline, we asked participants how often they got news from different sources over the past four weeks, including Facebook, cable TV, print, and radio news, borrowing a standard survey question from the Pew Research Center (2018). At endline, we again asked how often they got news from those same sources, “relative to what is typical for you.” For the participants who reported having a Twitter handle, we gathered data on number of tweets in the four weeks before baseline began and in the four weeks between midline and endline. This

allows a non-self-reported measure of one kind of potential substitution away from Facebook.<sup>13</sup>

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

*Political Engagement.*—We have two measures of political engagement. First, we measure whether participants voted in the 2018 midterm election, by matching participants on name, birth year, and zip code to a voting database supplied to Stanford by L2, a voting data provider. See online Appendix C for details on the match process. Second, we measure whether participants clicked on any of the links in the post-endline politics email.

*Political Polarization.*—There are a variety of ways to measure political polarization (see, for example, Gentzkow 2016), and we use both standard and novel measures. First, we included standard “feeling thermometer” questions capturing how “warm or cold” participants felt toward the Democratic 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. For this and all other polarization variables, we include independents who lean toward a party, and we drop independents who do not lean toward either party.

<sup>13</sup>In our pre-analysis plan, we grouped this *number of tweets* variables in the substitute news sources family, but one might also think of it as a “substitute time use” because Twitter is not only used to read news.

Second, the *Trump affective polarization* variable is the thermometer warmth toward President Trump for Republicans, and  $-1$  times the thermometer warmth toward President Trump for Democrats. 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 1 plus the length (in characters of text) of her response about the other party minus the natural log of 1 plus the length of her response about her own party. Fourth, we asked people how often they saw news that made them better understand the point of view of the Republican Party, and a parallel question for news about the Democratic Party. *Congenial news exposure* is the respondent's answer about her own political party minus her answer for the other party.

Fifth, we asked opinions about nine current political issues, such as "To what extent do you think that free trade agreements between the US and other countries have been a good thing or a bad thing for the United States?" These nine questions were all adapted from recent Pew Center and Gallup opinion polls. 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.<sup>14</sup> Finally, *vote polarization* measures the strength of preferences for the congressional candidate of the respondent's party in the midterm election.<sup>15</sup>

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

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

<sup>14</sup>Specifically, for each issue or belief question  $q$ , we normalize responses by the standard deviation in the Control group, determine Democrats' and Republicans' average responses  $\mu_q^D$  and  $\mu_q^R$ , recenter so that  $\mu_q^D + \mu_q^R = 0$ , and resign so that  $\mu_q^R > 0$ . Define  $\tilde{y}_{iq}$  as individual  $i$ 's normalized, recentered, and re-signed response to question  $q$ , multiplied by  $-1$  if  $i$  is a Democrat. Thus  $\tilde{y}_{iq}$  reflects the strength of individual  $i$ 's agreement with the average view of her party instead of the other party. For *issue polarization*, further define  $\sigma_q$  as the Control group within-person standard deviation of  $\tilde{y}_{iq}$  for question  $q$ . This measures how much people's views change between baseline and end-line, 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  $Y_i = \sum_q \tilde{y}_{iq} \sigma_q$ . Online Appendix Table A15 shows that the *issue polarization* results are nearly identical if we set  $\sigma_q = 1$ .

<sup>15</sup>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.



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 (2014) and Stone and Shiffman (1994). The variable *SMS happiness* is the answer to the question, “Overall, how happy do you feel right now on a scale from 1 (not at all happy) to 10 (completely happy)?” The variable *SMS positive emotion* is an indicator variable for whether the participant reports a positive emotion when asked, “What best describes how you felt over the last ten minutes?” with possible responses such as “angry,” “worried,” “loving/tender,” etc. Finally, *SMS not lonely* uses the answer to the question, “How lonely are you feeling right now on a scale from 1 (not at all lonely) to 10 (very lonely)?”

*Post-Experiment Facebook Use.*—We have four measures of planned and actual post-experiment Facebook use. First, *planned post-study use change* is the extent to which participants plan to use Facebook more or less than they had before they started the study. (This was included only in the endline survey.) Second, *clicked time limit email* is an indicator for whether the respondent clicked any of the links in the post-endline social media time limit email. Third, *speed of reactivation* is  $-1$  times the natural log of 1 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 17. Fourth, *Facebook mobile app use* is the natural log of 1 plus the number of minutes per day that the participant reported using Facebook on their phone in the post-endline survey.

*Opinions about Facebook.*—We asked eight questions eliciting people’s opinions about Facebook, such as “To what extent do you think Facebook is good or bad for society?” and “To what extent do you think Facebook makes people more or less politically polarized?” Each of these eight responses was on a ten-point scale. In the endline survey only, we also asked *Deactivation bad*: “As part of this study, you were asked to deactivate your Facebook account for [24 hours/4 weeks]. To what extent do you think that deactivating your account was good or bad for you?” Finally, we also included two open answer text boxes in which we asked people to write out the most important positive and negative impacts that Facebook has on their lives. The *positive impacts* and *negative impacts* variables are the natural log of 1 plus the count of characters in the respective text box.

*Secondary Outcomes.*—We also consider the following two outcomes, which we labeled as “secondary” in our pre-analysis plan. First, we consider the standard generic ballot question. At baseline, we asked “If the elections for US Congress were being held today, would you vote for the Republican Party’s candidate or the Democratic Party’s candidate for Congress in your district?” To increase precision,



TABLE 1—SAMPLE SIZES

Phase	Sample size
Recruitment and baseline	$N = 1,892,191$ were shown ads $N = 32,201$ clicked on ads $N = 22,324$ completed pre-screen survey $N = 20,959$ were from United States and born between 1900 and 2000 $N = 17,335$ had $15 < \text{daily Facebook minutes} \leq 600$ $N = 7,455$ consented to participate $N = 3,910$ finished baseline $N = 2,897$ had valid baseline and were randomized, of which:
Midline	$N = 2,897$ began midline $N = 2,743$ received a price offer, of which: $N = 1,661$ were in impact evaluation sample
Endline	$N = 2,710$ began endline $N = 2,684$ finished endline, of which: $N = 1,637$ were in impact evaluation sample
Post-endline	$N = 2,067$ reported Facebook mobile app use, of which: $N = 1,219$ were in impact evaluation sample

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 asked people to report whether they had voted (at endline) and planned to vote (at baseline) in the 2018 midterm. We labeled this as secondary because it is superseded by the administrative voting data from L2.

We also gathered contributions to political campaigns from the Federal Election Commission database. In our pre-analysis plan, we labeled this as secondary because very few Americans contribute to political campaigns, and we did not expect to be able to detect effects from four weeks of deactivation. Indeed, only one person in the impact evaluation sample donated to a political party between the October 2018 midline survey and July 2019. As a result, we deviate from the pre-analysis plan by dropping this from our analysis.

## II. Descriptive Statistics

Table 1 shows sample sizes at each step of our experiment, from the 1.9 million Facebook users who were shown our ads, to the 1,661 subjects in the impact evaluation sample. Table 2 quantifies the representativeness of our sample on observables, by comparing the demographics of our impact evaluation sample to our estimate of the average demographics of adult Facebook users and to the US adult population. Comparing column 1 to columns 2 and 3, we see that our sample is relatively high-income, well-educated, female, young, and Democratic, and uses Facebook

TABLE 2—SAMPLE DEMOGRAPHICS

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

*Notes:* Column 1 presents average demographics for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered  $p = \$102$  or  $p = \$0$  to do so. Column 2 presents our estimate of average demographics of American adults with a Facebook account. The top five numbers in column 2 are inferred from a Pew Research Center (2018f) survey of social media use by demographic group. The bottom number in column 2 (the average of 45 minutes of Facebook use per day) is approximated on those basis of sources such as Facebook (2016) and Molla and Wagner (2018). Column 3 presents average demographics of American adults. The top five numbers are from the 2017 American Community Survey (US Census Bureau 2017), and the Republican and Democrat shares are from the 2016 American National Election Study (American National Election Studies 2016).

relatively heavily.<sup>16</sup> Online Appendix Table A14 shows that Treatment and Control are balanced on observables.

Table 3 documents very high response rates to the endline and post-endline surveys and subjective well-being text messages. Of the 580 people in the Treatment group, only 7 failed to complete the endline survey. Of the 1,081 people in the Control group, only 17 failed to complete endline. The average participant responded to 92 percent of daily text messages, well above the 75 percent required in order to receive the completion payment.<sup>17</sup> Treatment and Control have statistically equal response rates to the endline survey and subjective well-being text messages. A marginally significantly larger share of the Treatment group responded to the post-endline survey; this is less worrisome because Facebook mobile app use is the only variable from that survey for which we calculate treatment effects, and we show in online Appendix Table A13 that using Lee (2009) bounds to account for attrition does not change the conclusions. Finally, Table 3 also reports the high level of compliance with our deactivation treatment: treatment group participants were deactivated on 90 percent of checks between October 13 (the first day after the 24-hour post-midline deactivation period) and November 7 (the day before endline), against 2 percent for Control.

As described above, if Treatment group members were found to have active accounts, we sent an email informing them of this and asking them to promptly deactivate, along with a survey asking why they were not deactivated. From these

<sup>16</sup>In online Appendix Figures A17, A18, A19, and A20, we find that the two demographic variables that we prespecified as moderators, age and political party, do not appear to systematically moderate treatment effects. Furthermore, Figure 9 provides no systematic evidence that the effects vary for people who use Facebook more versus less heavily before baseline. This suggests that reweighting the sample for representativeness on these observables would not substantively change the estimated effects, although it would increase the standard errors.

<sup>17</sup>Online Appendix Figure A26 shows the text message response rate by day (response rates declined slightly over the course of the experiment) and shows that Treatment and Control response rates are statistically balanced in all days of the deactivation period.

TABLE 3—SURVEY RESPONSE AND TREATMENT COMPLIANCE RATES

Variable	Treatment mean/SD (1)	Control mean/SD (2)	<i>t</i> -test <i>p</i> -value (1) – (2)
Completed endline survey	0.99 (0.11)	0.98 (0.12)	0.54
Share of text messages completed	0.92 (0.20)	0.93 (0.18)	0.45
Completed post-endline survey	0.95 (0.23)	0.92 (0.26)	0.07
Share days deactivated	0.90 (0.29)	0.02 (0.13)	0.00
Observations	580	1,081	

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

surveys, along with email interactions and formal qualitative interviews following our summer 2018 pilot study, we conclude that most Treatment group members who did reactivate fall into one of two groups. The first group consists of a small number of users who changed their mind about participating in the experiment and reactivated intentionally. The second group consists of users who briefly reactivated by accident, for example because they logged in to another app or online service using their Facebook account credentials.

Online Appendix Figure A27 shows the cumulative distribution of the share of time deactivated for the Treatment group, and online Appendix Figure A28 shows the distribution of reasons for deactivation among those for whom this share was less than 1. Together, these figures suggest that the small group of intentional reactivators accounts for the vast majority of Treatment group noncompliance. Given this, combined with the fact that the Control group was also found to be deactivated for a small share of weeks 1–4, we will analyze the experiment as a randomized encouragement design.

### III. Empirical Strategy

#### A. Pre-Analysis Plan

We submitted our pre-analysis plan on October 12, as this was the final day before the Treatment and Control groups could have begun to differ. We submitted a slightly updated pre-analysis plan on November 7, the day before endline, with only one substantive change: on the basis of data on reasons for non-compliance described above, we specified that our primary specifications would use IV estimates instead of intent-to-treat estimates. The pre-analysis plan specified three things. First, it specified the outcome variables and families of outcome variables as described above, including which specific variables are included in the index for

each family and which outcomes are “secondary.” Versions of Figures 2, 3, 5, 6, 7, and 12 appear as figure shells in the pre-analysis plan, although we changed some variable labels as well as the order in which we present the families of outcome variables for expositional purposes. Second, the pre-analysis plan specified the moderators we use when testing for heterogeneous treatment effects, including which moderators are “secondary.” Third, it specified the two regression specifications and the estimation sample as described below.

### B. Empirical Strategy

To estimate the local average treatment effect (LATE) of Facebook deactivation, define  $Y_i$  as some outcome measured at endline, and  $\mathbf{Y}_i^b$  as a vector including the baseline value of the outcome and the baseline value of the index that includes the outcome.<sup>18</sup> Define  $D_i$  as the percent of deactivation checks between October 13 and November 7 that person  $i$  is observed to be deactivated. Define  $T_i \in \{1, 0\}$  as a Treatment group indicator, and  $\nu_s$  as the vector of the 48 stratum dummies. We estimate local average treatment effects of deactivation using the following regression:

$$(1) \quad Y_i = \tau D_i + \rho \mathbf{Y}_i^b + \nu_s + \varepsilon_i,$$

instrumenting for  $D_i$  with  $T_i$ . In equation (1),  $\tau$  measures the local average treatment effect of deactivation for people induced to deactivate by the promised \$102 payment.<sup>19</sup>

The base sample for all regressions is the “impact evaluation sample”; again, participants who were willing to accept less than \$102 to deactivate in weeks 1–4 (the four weeks after midline) and were offered  $p = \$102$  or  $p = \$0$  to do so. For the political polarization outcomes, the sample includes only Democrats and Republicans, as well as independents who lean toward one party or the other. Sample sizes sometimes differ across outcomes due to missing data: for example, the post-endline survey has higher non-response than the endline survey, and many participants do not have Twitter accounts.

We use robust standard errors in all regressions.

<sup>18</sup>  $\mathbf{Y}_i^b$  excludes the baseline value of the outcome for outcomes such as clicks on post-endline emails that do not have a baseline value.  $\mathbf{Y}_i^b$  excludes the baseline index when  $Y_i$  is not included in an index. When  $Y_i$  is an index,  $\mathbf{Y}_i^b$  is simply the baseline value of the index.

<sup>19</sup> 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:

$$(2) \quad Y_i = \tau D_i H_i + \beta H_i + \rho \mathbf{Y}_i^b + \nu_s + \varepsilon_i,$$

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

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

#### IV. Impact Evaluation

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

In the body of the paper, we present figures with local average treatment effects and 95 percent confidence intervals from estimates of equation (1), with outcome variables  $Y_i$  normalized so that the Control group standard deviation equals 1. Online Appendix Tables A10 and A11 provide numerical regression results for all individual outcome variables in both normalized (standard deviation) units, as in the figures, and unnormalized (original) units. Online Appendix Table A12 provides numerical regression results for all nine summary indices. These tables also provide unadjusted  $p$ -values and “sharpened” False Discovery Rate (FDR)-adjusted  $p$ -values following the procedure of Benjamini, Krieger, and Yekutieli (2006), as outlined by Anderson (2008). The unadjusted  $p$ -values are appropriate for readers with a priori interest in one specific outcome. The FDR-adjusted  $p$ -values for the individual outcomes limit the expected proportion of false rejections of null hypotheses across all individual outcomes reported in the paper, while the FDR-adjusted  $p$ -values for the indices limit the expected proportion of false rejections of null hypotheses across the nine indices. The sharpened FDR-adjusted  $p$ -values are less conservative than the unadjusted  $p$ -values for  $p$ -values greater than about 0.15, and more conservative for unadjusted  $p$ -values less than that.

##### A. Substitutes for Facebook

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

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

The first row confirms that the treatment indeed reduced Facebook use as intended. At endline, the Control group reported that they had used Facebook for an average of 59.53 minutes per day over the past four weeks, and the local average

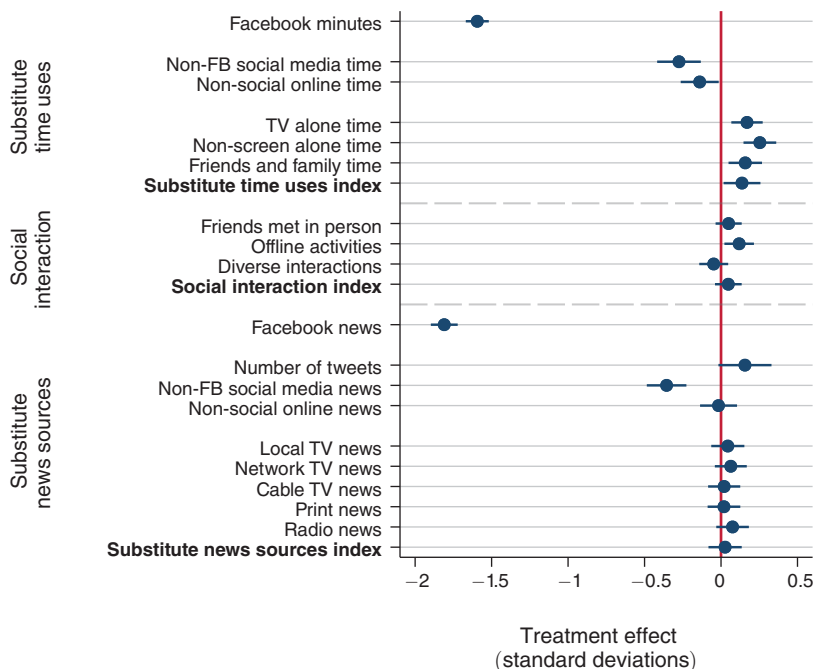


FIGURE 2. SUBSTITUTES FOR FACEBOOK

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

treatment effect of deactivation is 59.58 minutes per day.<sup>20</sup> As shown in Figure 2, this corresponds to a reduction of 1.59 standard deviations.

We find that Facebook deactivation *reduced* time devoted to other online activities. Time using non-Facebook social media falls by a quarter point on our five-point scale (0.27 SD), and time on non-social online activities falls by 0.12 points (0.14 SD). Thus, Facebook appears to be a complement rather than a substitute for other online activities. This makes sense to the extent that deactivating Facebook makes people less likely to be using their phones or computers in the first place, and less likely to follow Facebook links that direct to non-Facebook sites (e.g., a news website or Twitter post). Furthermore, the Treatment group may have avoided logging into other apps such as Spotify and Tinder because we had informed participants that using Facebook to actively log into other apps would reactivate Facebook.

Rows 4–7 of Figure 2 suggest that the 60 minutes freed up by not using Facebook, as well as the additional minutes from reductions in other online activities, were

<sup>20</sup>Online Appendix Table A5 reports baseline means of our time use variables. The mean of self-reported Facebook minutes at baseline is 74.5 minutes per day, and the mean of reported minutes using the Facebook mobile app at baseline is 60 minutes per day.



allocated to both solitary and social activities offline. Solitary television watching increases by 0.17 points on our scale (0.17 SD), other solitary offline activities increase by 0.23 points (0.25 SD), and time devoted to spending time with friends and family increases by 0.14 points (0.16 SD). The substitute time uses index, which does not include *Facebook minutes*, shows an increase in overall non-Facebook activities. All of the online and offline time use effects are highly significant with and without adjustment for multiple hypothesis testing.

The middle group of outcomes in Figure 2 contains measures of social interaction. Deactivation increased the count of offline activities that people reported doing at least once last week by about 0.18 (0.12 SD). Online Appendix Figure A29 shows that the specific activities with the largest point estimates are going out to dinner, getting together with friends, and spending time with parents. The point estimates for the other offline activities we measure (going to the cinema, talking to friends on the phone, going to a party, going shopping, and spending time with your kids) are all very close to zero. Notwithstanding the positive effects on *offline activities*, there are no statistically significant effects on the number of friends that participants listed as having met in person last week, or on *diverse interactions* (whether or not they interacted with someone who voted differently in the last presidential election or interacted with someone from another country). We find no effects on the social interaction index, although the point estimate is positive.

The bottom group of outcomes in Figure 2 measures news consumption. As with the substitute time uses, the endline survey asked participants how much time they spent getting news from each source in the last four weeks relative to what is typical for them; “Same” is again the average answer in the Control group. As expected, Facebook deactivation substantially reduced the extent to which people said they relied on Facebook as a news source. Consistent with the time use results, the Treatment group also got substantially less news from non-Facebook social media sites (0.36 SD). The point estimates for print, radio, and TV news are all positive but statistically insignificant. Facebook deactivation has a positive but insignificant effect on Twitter use. As we discuss below in the news knowledge results, deactivation reduced the total time subjects report spending consuming news by 8 minutes per day, or 15 percent of the Control group mean of 52 minutes.

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

### B. Effects on News and Political Outcomes

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

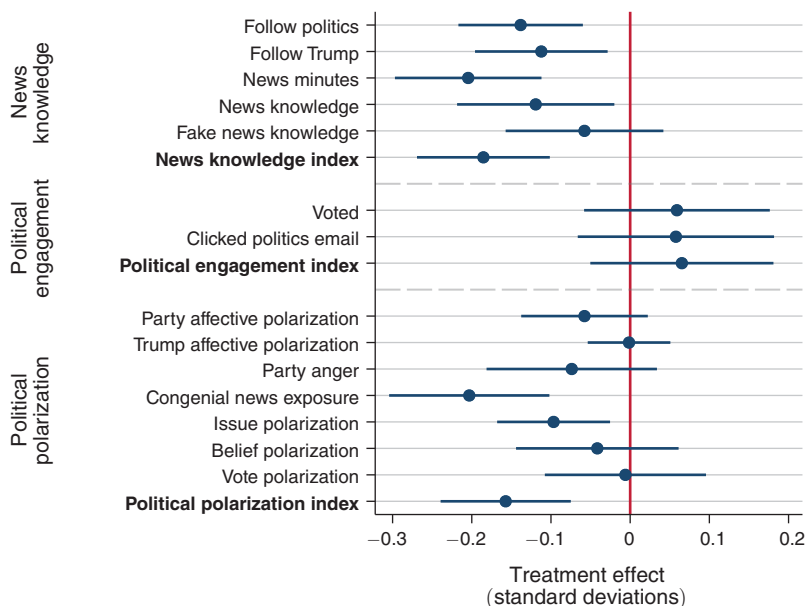


FIGURE 3. EFFECTS ON NEWS AND POLITICAL OUTCOMES

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

Deactivation caused substantial reductions in both self-reported attention to news and directly measured news knowledge. The top three rows show that deactivation reduced how much people reported they followed news about politics and about President Trump (by 0.14 and 0.11 SD, respectively), as well as the average minutes per day spent consuming news (a drop of 8 minutes per day, or 15 percent of the control group mean). Accuracy on our news knowledge quiz fell by 0.12 standard deviations.<sup>21</sup> Tangibly, the Control group answered an average of 7.26 out of the 10 news knowledge questions correctly (counting “unsure” as one-half correct), and deactivation reduced this average by 0.14. There is no detectable effect on fake news knowledge, possibly reflecting the limited reach of even the highly shared fake news items included in our survey. Overall, deactivation reduced the news knowledge index by about 0.19 standard deviations.

There are no statistically detectable effects on political engagement. As reported in online Appendix Tables A10 and A11, the point estimates suggest that deactivation increased turnout by three percentage points according to the administrative data and decreased turnout by three percentage points according to the self-reported

<sup>21</sup> Online Appendix G presents more analysis of the effects on news knowledge, including effects on each individual news knowledge and fake news knowledge question. All but one of the point estimates for the ten news knowledge questions is negative. The news knowledge questions with the largest effects involve correctly responding that Elizabeth Warren’s DNA test had revealed Native American ancestry and that Jeff Sessions had resigned at President Trump’s request. There was also a statistically significant difference in knowledge about one fake news story: the Treatment group was less likely to correctly respond that Cesar Sayoc, the suspect in an act of domestic terrorism directed at critics of President Trump, was not a registered Democrat.

data, and neither estimate is statistically different from zero. Similarly, the Treatment and Control groups are statistically equally likely to have clicked on any link in the post-endline politics email. Online Appendix Figure A35 does show a marginally significant negative effect on *voted Republican*, suggesting that deactivation may have reduced support for Republican congressional candidates. The unadjusted  $p$ -value is 0.06, the sharpened FDR-adjusted  $p$ -value is 0.08, and we had labeled this as a “secondary outcome” in our pre-analysis plan.

Prior research has shown that people tend to be exposed to ideologically congenial news content in general (Gentzkow and Shapiro 2011) and on Facebook in particular (Bakshy, Messing, and Adamic 2015). Thus, the finding above that deactivation reduced news exposure naturally suggests that deactivation might have also reduced political polarization.

Indeed, deactivation did reduce political polarization. Point estimates are negative for all polarization measures. The largest and most significant individual effect is on *congenial news exposure*: deactivation decreased the number of times that people reportedly saw news that made them better understand the point of view of their own political party relative to the other party. Deactivation also decreased issue polarization, which Fiorina and Abrams (2008) singles out as the “most direct” way of measuring polarization.<sup>22</sup> Online Appendix Table A10 shows that both of these effects are highly significant after adjusting for multiple hypothesis testing. The other measures with the largest point estimates are *party anger* and *party affective polarization*, although these individual effects are not statistically significant. Overall, deactivation reduced the political polarization index by about 0.16 standard deviations.<sup>23</sup>

Figure 4 illustrates how deactivation reduced issue polarization, by plotting the distribution of “issue opinions” for Democrats and Republicans in Treatment and Control at endline. Our *issue opinions* measure exactly parallels the *issue polarization* variable used in the regressions, except that we keep opinions on a left-to-right scale, with more negative indicating more agreement with the average Democratic opinion, and more positive indicating more agreement with the average Republican opinion. (By contrast, the issue polarization variable multiplies Democrats’ responses by  $-1$ , so that a more positive value reflects more agreement with the average opinion in one’s political party.) We then normalize *issue opinions* to have a standard deviation of 1 in the Control group. The figure shows that deactivation moves both Democrats and Republicans visibly toward the center. In the Control group, the issue opinions of the average Democrat and the average Republican differ by 1.47 standard deviations. In the Treatment group, this difference is 1.35 standard deviations: about 8 percent less.

Are these polarization effects large or small? As one benchmark, we can compare these effects to the increase in political polarization in the United States since 1996,

<sup>22</sup>Online Appendix Figure A30 presents results for each of the issue polarization questions. The issues for which deactivation caused the largest decrease in polarization were the direction of racial bias in policing and whether the Mueller investigation is biased.

<sup>23</sup>Like all of our outcome families, the polarization index includes a range of different outcomes with different interpretations. Exposure to congenial news is conceptually different from affective polarization and issue polarization. Online Appendix Table A16 shows that the effect on the political polarization index is robust to excluding each of the seven individual component variables in turn, although the point estimate moves toward zero and the unadjusted  $p$ -value rises to 0.09 when omitting *congenial news exposure*.

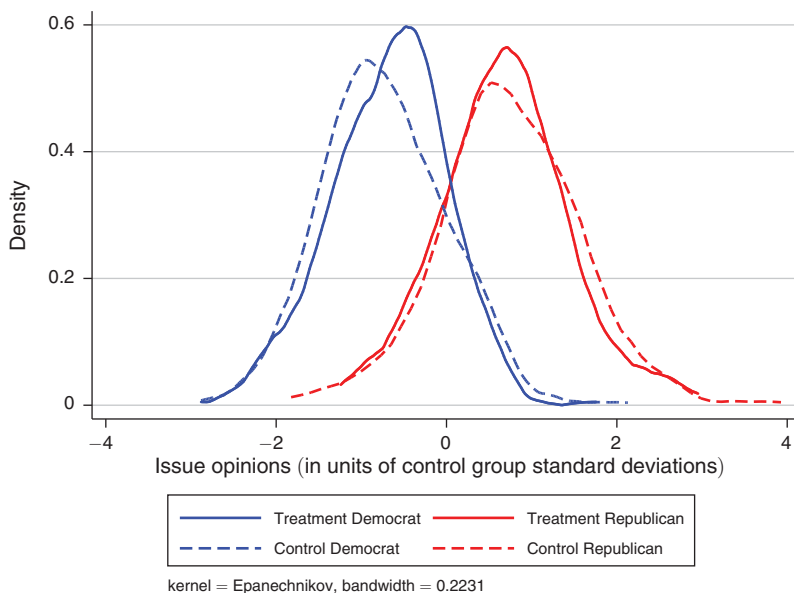


FIGURE 4. ISSUE OPINIONS BY PARTY AT ENDLINE

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

well before the advent of social media. Using data from the American National Election Studies, Boxell (2018) calculates that the change in a different index of polarization measures increased by 0.38 standard deviations between 1996 and 2016. The 0.16 standard deviation effect of Facebook deactivation on political polarization in our sample is about 42 percent as large as this increase.<sup>24</sup>

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

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

### C. Effects on Subjective Well-Being

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

We find that deactivation indeed significantly increases SWB. All but one of the ten point estimates are positive. The magnitudes are relatively small overall, with the largest and most significant effects on *life satisfaction* (0.12 SD), *anxiety* (0.10 SD), *depression* (0.09 SD), and *happiness* (0.08 SD).<sup>25</sup> All of these effects remain significant after adjusting for multiple hypothesis testing. The text message based measures of happiness are not significantly different from zero, with positive point estimates ranging from 0.01 SD to 0.06 SD. Deactivation improved our overall SWB index by 0.09 standard deviations.

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

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

As a third benchmark, online Appendix Table A17 presents a regression of our baseline SWB index on key demographics (income, college completion, gender, race, age, and political party). College completion is conditionally associated with 0.23 standard deviations higher SWB. Thus, the effect of deactivating Facebook is just over one-third of the conditional difference in subjective well-being between college graduates and everyone else. The table also shows that

<sup>25</sup> Online Appendix Figure A34 presents results for the individual questions within the *happiness*, *life satisfaction*, and *loneliness* scales.

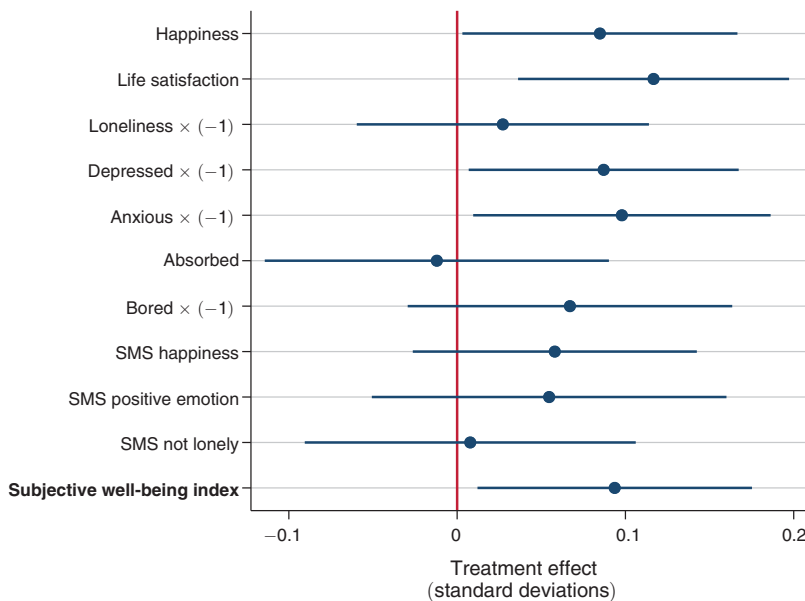


FIGURE 5. EFFECTS ON SUBJECTIVE WELL-BEING

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

a \$10,000 increase in income is conditionally associated with a 0.027 standard deviation increase in SWB. Thus, the effect of deactivating Facebook is equal to the conditional difference in subjective well-being from about \$30,000 additional income. This income equivalent is large because “money doesn’t buy happiness”: although income is correlated with SWB, the slope of that relationship is not very steep.

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

We can also compare our SWB effects to what we would have estimated using the kind of correlational approach taken by many previous non-experimental studies. These studies often have specific designs and outcomes that don’t map closely to our paper, so it is difficult to directly compare effect sizes with other papers. We can, however, replicate the empirical strategy of simple correlation studies in our data, and compare our cross-sectional correlations to the experimental results. To do this, we regress SWB outcomes at baseline on daily average Facebook use over the past four weeks as of baseline, divided by the local average treatment effect 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.<sup>26</sup>

The baseline correlation between our SWB index and Facebook use is about three times larger than the experimental estimate of the treatment effect of deactivation (about 0.23 SD compared to 0.09 SD), and the point estimates are highly statistically significantly different. Controlling for basic demographics brings down the non-experimental estimate somewhat, but it remains economically and statistically larger than our experimental estimate. Online Appendix Figure A32 presents the full results for all SWB outcomes.<sup>27</sup> These findings are consistent with reverse causality, for example if people who are lonely or depressed spending more time on Facebook, or with omitted variables, for example if lower socioeconomic status is associated with both heavy use and lower well-being. They could also reflect a difference between the relatively short-term effects measured in our experiment and the longer-term effects picked up in the cross section. However, the lack of a detectable trend in treatment effects on the text message outcomes over the course of our experiment (as noted above and seen in online Appendix Figure A31) points away from this hypothesis.

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

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

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

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

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

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

<sup>26</sup>Specifically, the non-experimental estimates are from the following regression:

$$(3) \quad Y_i^b = \tau \tilde{H}_i + \beta \mathbf{X}_i + \epsilon_i,$$

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

<sup>27</sup>One could also do similar experimental versus non-experimental comparisons for other outcomes, but we have done this only for SWB because SWB is the focus of the non-experimental literature in this area.

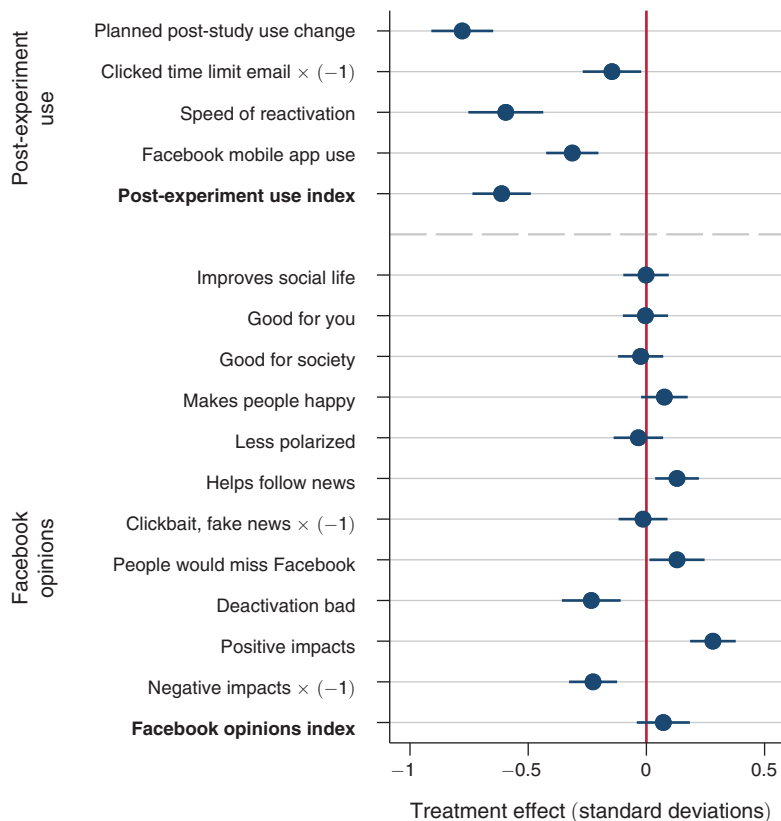


FIGURE 6. EFFECTS ON POST-EXPERIMENT FACEBOOK USE AND OPINIONS

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

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

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

#### D. Post-Experiment Facebook Use and Opinions

Figure 6 presents effects of deactivation on post-experiment demand for Facebook as well as participants' subjective opinions about Facebook. These results are closely related to the findings on subjective well-being, as we might expect participants who found deactivation increased their happiness would choose to use Facebook less in the future. They also speak more directly to the popular debate over whether social media are addictive and harmful. If deactivation

reduces post-experiment Facebook use, this is consistent with standard habit formation models such as Becker and Murphy (1988), or with learning models in which experiencing deactivation caused people to learn that they would be better off if they used Facebook less.<sup>28</sup>

Deactivation clearly reduced post-experiment demand for Facebook. These effects are very stark, with by far the largest magnitude of any of our main findings. The effect on reported intentions to use Facebook as of the endline survey is a reduction of 0.78 standard deviations: while the average Control group participant planned to reduce future Facebook use by 22 percent, deactivation caused the Treatment group to plan to reduce Facebook use by an additional 21 percent relative to Control. In our post-endline survey a month after the experiment ended, we measured whether people actually followed through on these intentions, by asking people how much time they had spent on the Facebook mobile app on the average day in the past week. Deactivation reduces this post-endline Facebook mobile app use by 12 minutes per day, or 0.31 standard deviations. This is a 23 percent reduction relative to the Control group mean of 53 minutes per day, lining up almost exactly with the planned reductions reported at endline. However, online Appendix Table A13 shows that the reduction is less than half as large (8 percent of the Control group mean) and not statistically significant (with a *t*-statistic of  $-1.16$ ) if we limit the sample to iPhone users who reported their usage as recorded by their Settings app, thereby excluding participants who were reporting personal estimates of their usage.

As a different (and non-self-reported) measure of post-experiment use, we can look at the speed with which people reactivated their Facebook accounts following the 24-hour post-endline period in which both Control and Treatment were deactivated. Figure 7 presents the share of our deactivation checks in which the Treatment and Control groups were deactivated, by day of the experiment.<sup>29</sup> By day 35, one week after the end of the experiment, 11 percent of the Treatment group was still deactivated, compared to 3 percent of the Control group. By day 91, nine weeks after the end of the experiment, 5 percent of the Treatment group was still deactivated, against 2.5 percent of Control. As Figure 6 shows, the local average treatment effect on the speed of reactivation is a highly significant 0.59 standard deviations. Overall, deactivation clearly decreased post-experiment use, reducing the index by 0.61 standard deviations. As introduced above, this is consistent with models of habit formation or learning.

The bottom group of outcomes in Figure 6 supplement the post-experiment use outcomes by measuring participants' qualitative opinions about Facebook. These are re-signed so that more positive means more positive opinions, so agreement with the statement that "Facebook exposes people to clickbait or false news stories" and the length of text about Facebook's negative impacts are both multiplied by  $(-1)$ .

<sup>28</sup> Online Appendix Figure A33 presents histograms of participants' opinions about Facebook at baseline. People are evenly divided on whether Facebook is good or bad for themselves and for society and whether Facebook makes people more or less happy. Consistent with our results, people tend to think that Facebook helps people to follow the news better and makes people more politically polarized.

<sup>29</sup> There is a slight dip in deactivation rates for the Treatment group seven days after the deactivation period began. This was caused by the fact that some participants failed to turn off a default setting in which Facebook reactivates users' profiles after seven days of deactivation. For technical reasons, our deactivation checking algorithm checked the entire Control group once every few days between midline and endline in order to check the Treatment group four times per day. After endline, we returned to checking all participants approximately once per day.

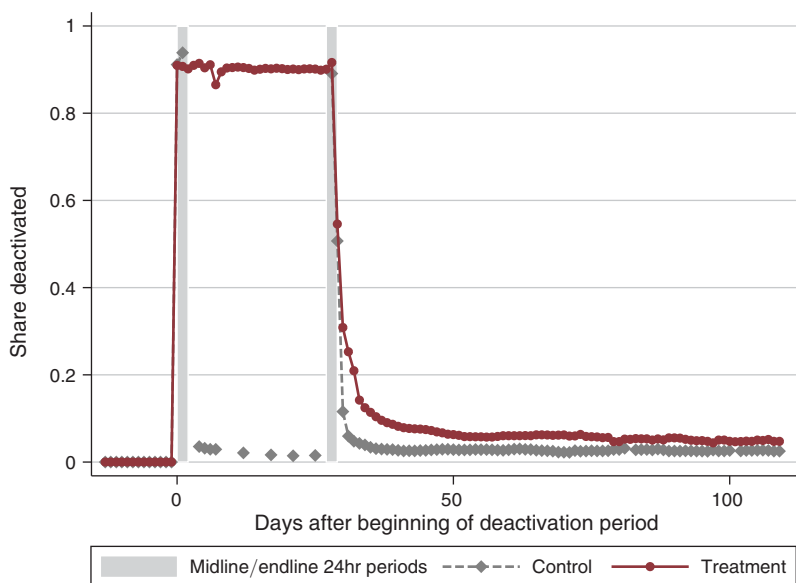


FIGURE 7. PROBABILITY OF BEING DEACTIVATED

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

The results are mixed. Deactivation increases the extent to which participants think Facebook helps them follow the news better, and it also makes participants agree more that people would miss Facebook if they stopped using it. On the other hand, participants who deactivated for four weeks instead of 24 hours were more likely to say that their deactivation was good for them.<sup>30</sup> Deactivation increases both the *positive impacts* and *negative impacts* variables, i.e., it makes people write more about both positive and negative aspects of Facebook. Overall, deactivation had no statistically significant effect on the Facebook opinions index.

Figure 8 presents the distributions of Treatment and Control responses to two key questions reflecting opinions about Facebook. Both Treatment and Control tended to agree that “if people spent less time on Facebook, they would soon realize that they don’t miss it,” but deactivation weakened that view. On this figure, the Treatment group’s average response on the scale from  $-5$  to  $+5$  was  $-1.8$ , while the Control group’s average response is  $-2.0$ . The right panel shows that both Treatment and Control tended to think that deactivation was good for them, but the Treatment group is more likely to think that their (longer) deactivation was good for them. On this figure, the Treatment group’s average response on the scale from

<sup>30</sup>One should be cautious in interpreting this effect, as it could result both from a change of opinion about Facebook and from the difference in length of the deactivation they were evaluating. As we shall see below, the Control group also tends to believe that deactivation was good for them, but the modal answer was 0 (i.e., neither good nor bad), suggesting that many people were indifferent to such a short deactivation.

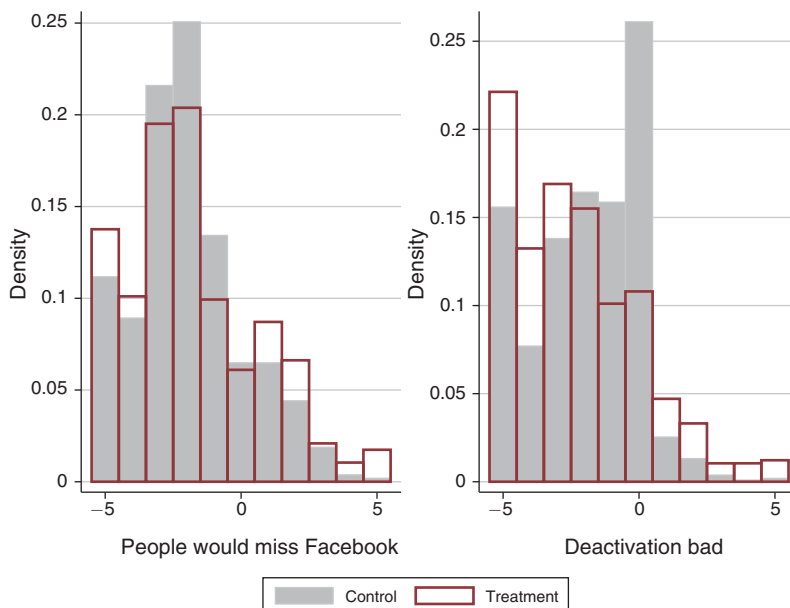


FIGURE 8. KEY OPINIONS ABOUT FACEBOOK IN TREATMENT AND CONTROL

*Notes:* This figure presents the distribution of responses in Treatment and Control for two key measures of opinions about Facebook. See Section IC for variable definitions.

−5 to +5 is −2.3, while the Control group’s average response is −1.9. Remarkably, about 80 percent of the Treatment group thought that deactivation was at least somewhat good for them, and the modal response was the strongest possible agreement that deactivation was good (the left-most bar on the histogram). In both panels, the Treatment group has a wider dispersion of responses, with more people strongly agreeing *and* more people strongly disagreeing. This highlights the importance of testing for treatment effect heterogeneity, which we will do in the next section.

To give a richer sense of how deactivation affected Facebook use, the post-endline survey included a free-response question in which we asked people to write how they had changed their Facebook use since participating in the study. We then use standard text analysis tools to determine how the Treatment and Control groups responded differently. Specifically, we processed the text by stemming words to their linguistic roots (for example, “changes,” “changing,” and “changed” all become “chang”), removing common “stop words” (such as “the” and “that”), and making lists of all one-, two-, three-, and four-word phrases that appeared five or more times in the sample. We then constructed Pearson’s  $\chi^2$  statistic, which measures the extent of differential usage rates between Treatment and Control; the phrases with the highest  $\chi^2$  are especially unbalanced between the two groups. This parallels Gentzkow and Shapiro’s (2011) approach to determining which phrases are used more by Republicans versus Democrats, except we determine which phrases are used more by Treatment versus Control.

The two panels of Table 4 present the 20 highest- $\chi^2$  phrases that were more common in Treatment and in Control. The Treatment group was relatively likely

TABLE 4—MOST COMMON DESCRIPTIONS OF FACEBOOK USE CHANGES

Phrases used more often by Treatment			Phrases used more often by Control		
Phrase	% Treatment	% Control	Phrase	% Treatment	% Control
Not use facebook anymor	0.90	0	Ha not chang	6.63	16.76
Not spend much time	1.08	0.36	Not chang sinc particip	0	0.99
Spend less time facebook	0.90	0.27	Ha not chang sinc	0.18	1.53
Have not use facebook	0.72	0.18	Chang sinc particip studi	0	0.81
Not use facebook much	0.72	0.18	Way use facebook ha	0.18	1.35
Spend lot less time	0.72	0.27	Usag ha not chang	0	0.72
Use much less	2.87	0.63	Chang way use facebook	0.18	1.26
Definit use facebook	0.54	0.18	Not chang	7.17	18.65
Use facebook lot less	0.54	0.18	Awar much time spend	0	0.63
Use facebook much less	0.54	0.18	Ha not	8.24	19.64
Not use facebook	3.05	1.17	Not much ha chang	0	0.54
Use littl bit less	0.54	0.27	Way use facebook	0.54	2.70
Have not use	1.25	0.18	Not think chang much	0	0.45
Ha not chang use	0.72	0.45	Not chang much use	0	0.45
Use facebook anymor	0.90	0.09	Use facebook slightli less	0	0.45
Think use less	1.61	0.45	More awar much time	0.18	0.99
No ha not chang	0.54	0.36	Chang sinc particip	0	1.08
Use news app	0.72	0.09	Much time spend	0.18	1.53
Still have not	0.90	0.18	Facebook ha not chang	0.72	2.07
Much less	4.84	1.17	Use slightli less	0	0.90

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

$$\chi^2 = \frac{(f_{p|T}f_{\sim p|C} - f_{p|C}f_{\sim p|T})^2}{(f_{p|T} + f_{p|C})(f_{p|T} + f_{\sim p|T})(f_{p|C} + f_{\sim p|C})(f_{\sim p|T} + f_{\sim p|C})}.$$

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

to write that they were using Facebook less or not at all (“use much less,” “not use facebook anymor,” “stop use facebook”) or more judiciously: the phrase “use news app” is mostly from people saying that they have switched to getting news from their phone’s news app instead of Facebook. By contrast, while a few of the Control group’s most common phrases indicate lower use (variants of “more aware much time spend” and “use facebook slightli less”), the great majority of their relatively common phrases indicate that their Facebook use has not changed.

To more deeply understand the ways in which deactivation changed people’s relationship to Facebook, we partnered with a team of qualitative researchers who analyzed our survey data and additional participant interviews (Baym, Wagman, and Persaud forthcoming). They find that many participants emphasized that their time off of Facebook led them to use the platform more “consciously,” aligning their behavior with their desired use. For example, some participants discussed avoiding their news feed and only looking at their Facebook groups, while others removed the Facebook app from their phones and only accessed the site using their computers.



### E. Heterogeneous Treatment Effects

*Individual Moderators.*—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* versus *light users*: that is, people whose baseline reported Facebook use was above versus below median. There is no consistent evidence that the effects are different for people who report being heavier users, perhaps because Facebook use is measured with noise.

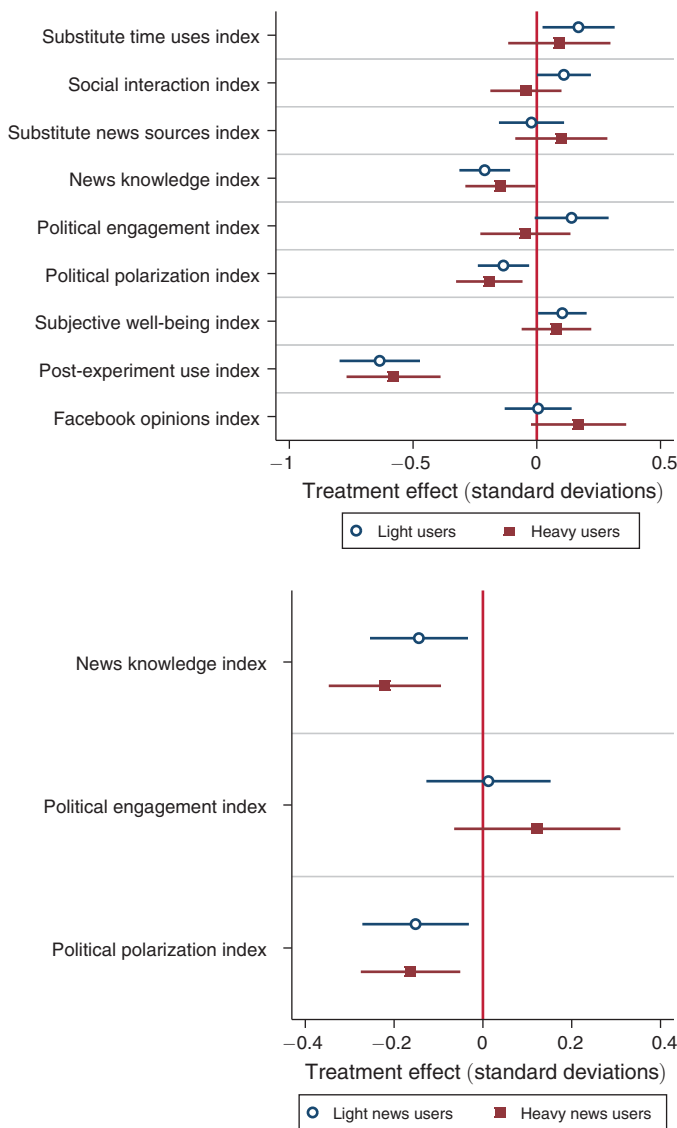
The second panel presents estimates for *heavy news users* versus *light news users*: that is, those who get news from Facebook fairly often or very often versus never, hardly ever, or sometimes. As one might expect, the estimated effects for news knowledge are larger for people who get more news from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

The third panel presents separate estimates for *active users* versus *passive users*. We measure this using two questions: share of active versus 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 versus passive users are defined as having above- versus below-median sum of their two responses to these questions. This moderator is of interest because of a set of papers cited in the introduction suggesting that passive Facebook use can be harmful to subjective well-being, while active use might be neutral or beneficial. Perhaps surprisingly, we see no differences in the effects of deactivation on the subjective well-being index. The pre-analysis plan specified that we would limit these tests to the four families reported in the figure.

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

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

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



(Continued)

FIGURE 9. HETEROGENEOUS TREATMENT EFFECTS

indistinguishable for four indices, and opposite-signed for the final index. This provides some support for the view that effect sizes would not be systematically different in the full Facebook user population including users with higher valuations.

Online Appendix Figure A9 presents one additional test of external validity that was suggested by a referee after the pre-analysis plan was submitted. We construct sample weights that match the impact evaluation sample to the observable characteristics of Facebook users in Table 2. Online Appendix Figure A9 shows that participants with below- versus above-median sample weights, that is, the types of

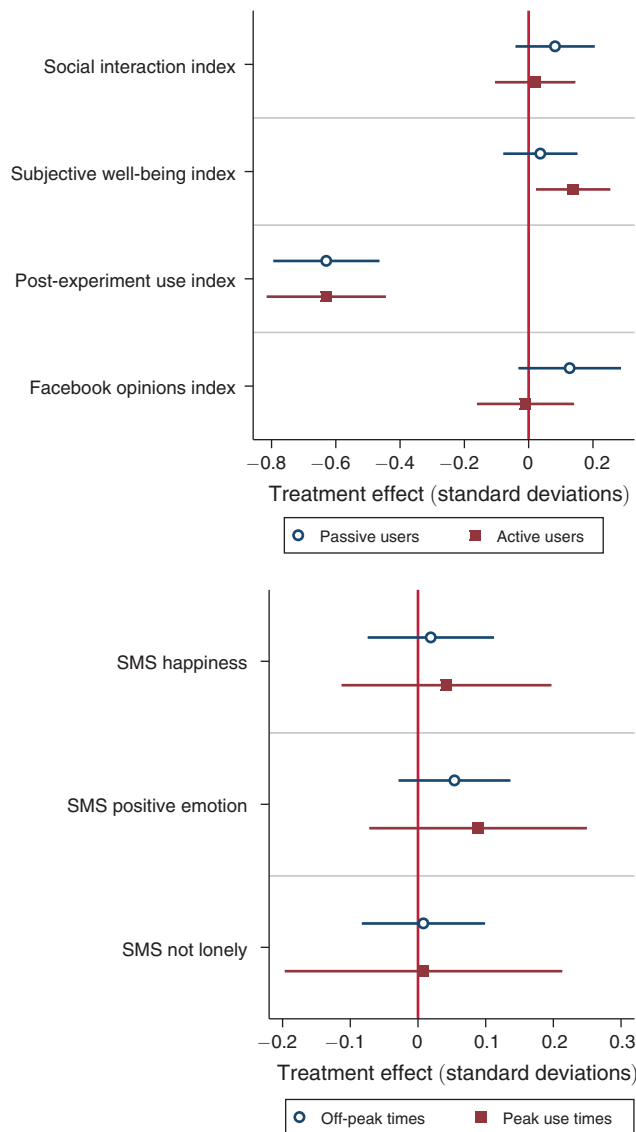


FIGURE 9. HETEROGENEOUS TREATMENT EFFECTS (CONTINUED)

*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 1. Error bars reflect 95 percent confidence intervals. See Section IC for variable definitions.

people who were especially likely versus unlikely to participate in the study, do not have systematically different treatment effects. This provides some further support for the view that effect sizes would be similar in the full Facebook user population.

Online Appendix F presents heterogeneous treatment effects on each individual outcome.

*All Possible Moderators.*—Many factors other than the specific variables we specified above might moderate treatment effects of Facebook deactivation. To search for additional possible moderators, we test whether any of the demographics or outcome variable indices collected at baseline might moderate treatment effects on the key outcomes of interest. We consider six outcomes: the latter five indices (news knowledge, political polarization, subjective well-being, post-experiment use, and Facebook opinions) plus the variable *Deactivation bad*, which we add because of the heterogeneity displayed in Figure 8. We consider 13 potential moderators: all 6 demographic variables listed in Table 2 (income, years of education, gender, race, age, and political party affiliation, which is on a seven-point scale from strongly Democratic to strongly Republican) and the baseline values of all 7 relevant indices.<sup>31</sup> We normalize each potential moderator to have a standard deviation of 1, and we denote normalized moderator  $k$  by  $X_i^k$ .

For all outcomes other than *Deactivation bad*, we estimate the following modified version of equation (1):

$$(4) \quad Y_i = \tau D_i + \alpha^k D_i X_i^k + \zeta X_i^k + \rho \mathbf{Y}_i^b + \nu_s + \varepsilon_i,$$

instrumenting for  $D_i$  and  $D_i X_i^k$  with  $T_i$  and  $T_i X_i^k$ . For *Deactivation bad*, we simply estimate  $Y_i = \alpha^k X_i^k + \varepsilon_i$  in the Treatment group only; this identifies what types of people in the Treatment group thought that deactivation was particularly good or bad. In total, we carry out 78 tests in 78 separate regressions: 13 potential moderators for each of the 6 outcomes.

There are many ways to estimate heterogeneous treatment effects, including causal forests (Athey, Tibshirani, and Wagner 2019) and lasso procedures. We chose this approach because it delivers easily interpretable estimates.

Figure 10 presents the interaction coefficients  $\hat{\alpha}^k$  and 95 percent confidence intervals for each of the six outcomes. To keep the figures concise, we plot only the five moderators with the largest absolute values of  $\hat{\alpha}^k$ , so there are another eight smaller unreported  $\hat{\alpha}^k$  coefficients for each outcome.

We highlight three key results. First, deactivation may reduce polarization more (i.e., Facebook use may increase polarization more) for older people, white people, and men. Second, Facebook deactivation has less positive effect on subjective well-being for people who have more offline social interactions and are already more happy at baseline. This suggests that Facebook use may have the unfortunate effect of reducing SWB more for people with greater social and psychological need. In our sample, these “higher-need” people also use Facebook more heavily. Third, people may have some intuition about whether they will like deactivation: people with more positive baseline opinions about Facebook are less likely to decrease their post-experiment use and less likely to think that deactivation was good for them.

<sup>31</sup> There are originally nine indices. We exclude the baseline substitute time uses index because it is not easily interpretable, and we exclude the baseline post-experiment use index because this only includes *Facebook mobile app use*.

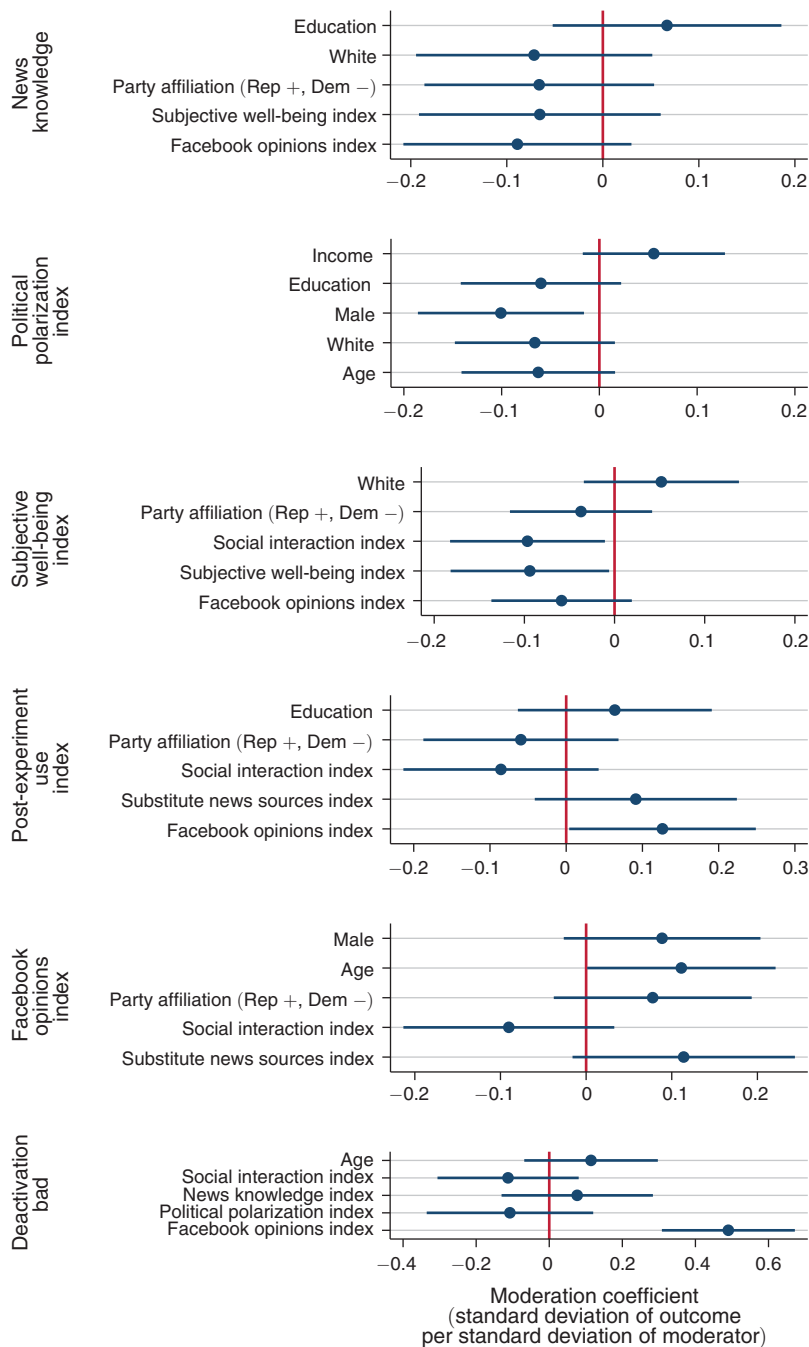


FIGURE 10. HETEROGENEOUS TREATMENT EFFECTS FOR ALL MODERATORS

*Notes:* This figure presents the moderators of local average treatment effects of Facebook deactivation estimated using equation (4). For each of the six outcomes, we present the five moderators with the largest moderation coefficients  $\hat{\alpha}^k$ . All outcome variables are normalized so that the Control group endline distribution has a standard deviation of 1, and all moderators are also normalized to have a standard deviation of 1. Error bars reflect 95 percent confidence intervals. See Section IC for variable definitions.

TABLE 5—PERCEIVED RESEARCHER AGENDA IN TREATMENT AND CONTROL

Variable	Treatment mean/SD (1)	Control mean/SD (2)	<i>t</i> -test <i>p</i> -value (1) – (2)
I don't think they had a particular agenda	0.43 (0.49)	0.44 (0.50)	0.59
Yes, wanted to show that Facebook is good for people	0.03 (0.18)	0.04 (0.19)	0.79
Yes, wanted to show that Facebook is bad for people	0.35 (0.48)	0.35 (0.48)	0.79
I am not sure	0.19 (0.39)	0.18 (0.38)	0.62
Observations	573	1,064	

*Notes:* The endline survey asked, “Do you think the researchers in this study had an agenda?” Columns 1 and 2 present the share of the Treatment and Control groups who gave each possible response. Column 3 presents *p*-values of tests of differences in means between the two groups.

### F. Experimenter Demand Effects

Most of our outcomes are self-reported, and it would have been difficult to further conceal the intent of the randomized experiment. This raises the possibility of experimenter demand effects, i.e., that survey responses depend on what participants think the researchers want them to say. To test for demand effects, the endline survey asked, “Do you think the researchers in this study had an agenda?” Table 5 presents the possible responses and shares by treatment group.

For demand effects to arise, participants must believe that the researchers want a particular pattern of responses. Table 5 shows that 62 percent of both Treatment and Control groups thought we had no particular agenda or were not sure. This suggests that demand effects would not arise for a solid majority of our sample. However, demand effects could arise for the remaining 38 percent.

For experimenter demand effects to bias our treatment effects, either (i) the Treatment and Control groups must have different beliefs about what the researchers want, or (ii) participants must sense what treatment group they are in and change their answers to generate the treatment effect that they think the researchers want (or don't want). Table 5 shows that possibility (i) is not true: perceived researcher agenda is closely balanced between Treatment and Control. To test for possibility (ii), we can estimate treatment effects separately for the subsample that thought that we “wanted to show that Facebook is bad for people” versus all other participants. If (ii) is true, then our results should be different in these two subsamples. Online Appendix Figure A36 shows that this is not the case: the effects on outcome indices that look “good” or “bad” for Facebook (e.g., news knowledge, political polarization, subjective well-being, and post-experiment use) are not statistically different, and there is no pattern of point estimates to suggest that the results are generally more “good” or “bad” in one of the two subsamples.

Of course, these tests are only suggestive. But combined with the fact that the non-self-reported outcomes paint a similar picture to the self-reports, these tests suggest that demand effects are unlikely to be a major source of bias in our results.



## V. Measuring the Consumer Surplus from Facebook

Quantifying the economic gains from free online services such as search and media is particularly important given that these services represent an increasingly large share of the global economy. This measurement has been particularly challenging because the lack of price variation (or any price at all) makes it impossible to use standard demand estimation to measure consumer surplus.<sup>32</sup> In this section, we present two back-of-the-envelope consumer surplus calculations. First, we employ the standard assumption that willingness-to-accept identifies consumer surplus. Second, we adjust consumer surplus to account for the possibility that deactivation might help people learn their true valuation of Facebook. This adjustment highlights the challenges in using willingness-to-accept as a measure of consumer welfare.

### A. Standard Consumer Surplus Estimate

In a standard model, willingness-to-accept to abstain from Facebook equals consumer surplus. Figure 11 presents the histogram of WTA to deactivate Facebook for the four weeks after midline instead of only the 24 hours after midline. The median is \$100, and almost 20 percent had valuations greater than \$500. After winsorizing valuations at \$1,000, the mean is \$203. After reweighting the sample to match the observable characteristics of Facebook users in Table 2, the median is still \$100, and the winsorized mean is \$180. Multiplying the mean by the estimated 172 million US Facebook users would imply that 27 days of Facebook generates \$31 billion of consumer surplus.

Our sample's WTA for Facebook abstention is larger than in most other studies, but not all. In an online panel weighted for national representativeness, Brynjolfsson, Eggers, and Gannamaneni (2018) estimates that the mean WTA to not use Facebook for one month is \$48, and that the median WTA to hypothetically stop using social media for one year was \$205 in 2016 and \$322 in 2017. In their sample of European college students, Brynjolfsson, Eggers, and Gannamaneni (2018) finds a median WTA of \$175 for one month.<sup>33</sup> In samples of college students, residents of a college town, and Amazon MTurk workers, Corrigan et al. (2018) estimates that the mean annualized WTA to deactivate Facebook ranges from \$1,139 to \$1,921, depending on the sample and the length of deactivation. In a sample of college students, Mosquera et al. (2018) estimates that the median (mean) WTA to not use Facebook for one week is \$15 (\$25). In an unincentivized (stated preference) survey of MTurk workers, Sunstein (forthcoming) found a \$1 per month median willingness-to-pay for Facebook and a \$59 per month median willingness-to-accept to not use Facebook.

There are many caveats to using this type of stylized calculation to approximate the consumer surplus from Facebook. First, we (and Corrigan et al.) required participants to deactivate their Facebook accounts instead of simply abstaining from logging in. For people who planned to avoid using other apps with Facebook logins

<sup>32</sup> As mentioned in the introduction, see Brynjolfsson and Saunders (2009); Byrne, Fernald, and Reinsdorf (2016); Nakamura, Samuels, and Soloveichik (2016); Brynjolfsson, Rock, and Syverson (2019); and Syverson (2017).

<sup>33</sup> Online Appendix Figure A37 compares our demand curve to the Brynjolfsson, Eggers, and Gannamaneni (2018) demand curves.

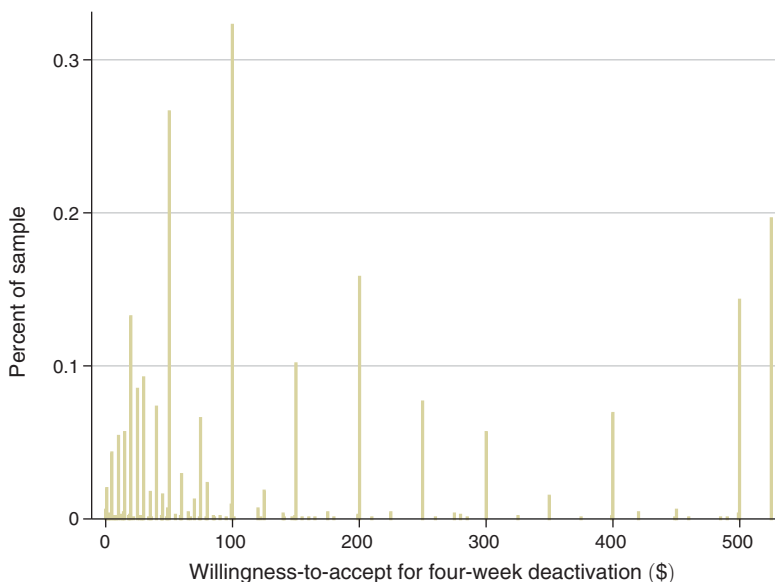


FIGURE 11. DISTRIBUTION OF WILLINGNESS-TO-ACCEPT TO DEACTIVATE FACEBOOK AFTER MIDLINE

*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.

in order to avoid reactivating their Facebook accounts, WTA overstates the value of Facebook access. Second, participants must believe the experimenter will in fact enforce deactivation; WTA could naturally be lower for a partially enforced or unenforced deactivation compared to an enforced deactivation. In some other studies, the method of enforcement was either not made clear *ex ante*, or enforcement was not fully carried out *ex post*.<sup>34</sup> Third, any survey sample is unlikely to be representative of the Facebook user population on both observable and unobservable characteristics. For example, we screened out people who reported using Facebook 15 minutes or less per day, and while we reweight the average WTAs to match the average observables of Facebook users (including average daily usage), this reweighting may implicitly overstate the WTA of people who don't use Facebook very much. Fourth, we (and all other existing studies) estimate people's Facebook valuations holding their networks fixed. Due to network externalities, valuations could be quite different if participants' friends and family also deactivated. Fifth, one should be careful in annualizing these estimates or comparing WTAs for different durations of abstention, as our study and several others find that the average per-day valuation varies with the duration. Sixth, as we will see, in practice people's WTA may not be closely held and could be easily anchored or manipulated, even in incentive

<sup>34</sup> Mosquera et al. told participants that they would "require" that they "not use their Facebook accounts" but did not give additional details. Brynjolfsson et al.'s WTA elicitation stated that the experimenters "will randomly pick 1 out of every 200 respondents and her/his selection will be fulfilled," and that they could enforce deactivation by observing subjects' time of last login, "given your permission." In practice, the deactivation was mostly not enforced: of the ten subjects randomly selected for enforcement, one gave permission.

compatible elicitations such as ours. Finally, this calculation fails to speak to the possibility that people misperceive Facebook's value. We turn to that issue now.

### *B. How Deactivation Affects Valuations*

It is often argued that social media users do not correctly perceive the ways in which social media could be addictive or make them unhappy. If this is the case, people's willingness-to-accept to abstain from Facebook would overstate "true" consumer surplus. For example, Alter (2018), Newport (2019), many popular media articles,<sup>35</sup> and organizations such as the Center for Humane Technology and Time to Log Off argue that Facebook and other digital technologies can be harmful and addictive. The Time to Log Off website argues that "everyone is spending too much time on their screens" and runs "digital detox campaigns." Sagioglu and Greitemeyer (2014) documents an "affective forecasting error": people predicted that spending 20 minutes on Facebook would make them feel better, but a treatment group randomly assigned to 20 minutes of Facebook browsing actually reported feeling worse.

Some of our results are also consistent with this argument. In the baseline survey, two-thirds of people agreed at least somewhat that "if people spent less time on Facebook, they would soon realize that they don't miss it." As reported earlier, about 80 percent of the Treatment group thought that deactivation was good for them, and both qualitative and quantitative data suggest that deactivation caused people to rethink and reoptimize their use.

The core of this argument is that people's social media use does not maximize their utility, and a "digital detox" might help them align social media demand with their own best interests. This idea is related to several existing economic models. In a model of projection bias (Loewenstein, O'Donoghue, and Rabin 2003), people might not correctly perceive that social media are habit forming or that their preferences might otherwise change after a "digital detox." In an experience good model, a "digital detox" might help consumers to learn their valuation of social media relative to other uses of time. Of course, both of these mechanisms could also affect demand after a period of deactivation, so it is not clear whether the WTA before deactivation or after deactivation is more normatively relevant.

To provide evidence on these issues, we elicited WTA at three points, as described earlier. First, on the midline survey, we elicited WTA to deactivate Facebook in "weeks 1–4" (the four weeks after midline). We call this WTA  $w_1$ . Second, just after telling people their BDM offer price on the midline survey, and thus whether they were expected to deactivate in weeks 1–4, we elicited WTA to deactivate in "weeks 5–8" (the four weeks after endline). We call this  $w_{2,1}$ . Third, on the endline survey, we elicited WTA to deactivate in weeks 5–8, after the Treatment group had experienced deactivation in weeks 1–4, but the Control group had not. We call this  $w_{2,2}$ .

<sup>35</sup> For example: Chris Ciaccia, "Facebook, Cocaine, Opioids: How Addictive Is the Social Network?" *Fox News*, December 29, 2017, <https://www.foxnews.com/tech/facebook-cocaine-opioids-how-addictive-is-the-social-network>; Will Oremus, "Addiction for Fun and Profit," *Slate*, November 10, 2017, <https://slate.com/technology/2017/11/facebook-was-designed-to-be-addictive-does-that-make-it-evil.html>.

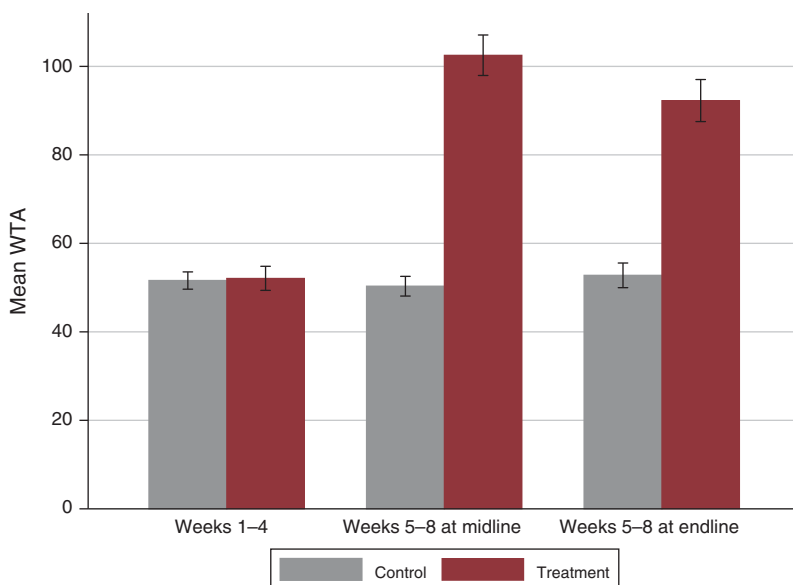


FIGURE 12. AVERAGE VALUATION OF FACEBOOK IN TREATMENT AND CONTROL

*Notes:* This figure presents the mean willingness-to-accept (WTA) to deactivate Facebook in Treatment and Control, for the impact evaluation sample: participants who were willing to accept less than \$102 to deactivate Facebook for the four weeks after midline and were offered  $p = \$102$  or  $p = \$0$  to do so. The first pair of bars is the mean WTA for deactivation in weeks 1–4, the four weeks after the midline survey. The second pair of bars is mean WTA for deactivation in weeks 5–8, the four weeks after the endline survey, as elicited in the midline survey. The third pair of bars is mean WTA for deactivation in weeks 5–8, as elicited in the endline survey.

The Control group's change in WTA for weeks 5–8,  $\Delta w_2 := w_{2,2} - w_{2,1}$ , captures any unpredicted time effect. The Treatment group's WTA change  $\Delta w_2$  reflects both the time effect and the unexpected change in valuation caused by deactivation. If the time effect is the same in both groups, then the difference-in-differences measures the effect of deactivation on valuations due to projection bias, learning, and similar mechanisms.

Figure 12 presents the average of WTA in Treatment and Control of  $w_1$ ,  $w_{2,1}$ , and  $w_{2,2}$ . Recall that the impact evaluation sample includes only people with  $w_1 < \$102$ , so these averages are less than the unconditional means discussed above and presented in Figure 11. Because of outliers in the WTAs for weeks 5–8, we must winsorize WTA. We winsorize at \$170 for this figure and our primary regression estimates, as this is the upper bound of the distribution of BDM offers that we actually made for deactivation.

The Treatment group's valuation for weeks 5–8 jumps substantially relative to its valuation for weeks 1–4, while the Control group's valuation for weeks 5–8 does not. We used open-answer questions in the post-endline survey and qualitative interviews to understand this change. Some of the large gap may be due to costs of deactivation being convex in the length of deactivation: some people in the Treatment group wrote that they were much less comfortable deactivating for eight weeks instead of four, as they would have to make much more extensive arrangements to communicate with friends, coworkers, and schoolmates during a longer deactivation. However, participants' open-answer responses suggest that the

Treatment group's WTA increase is also affected by anchoring on the \$102 BDM offer that was revealed after the elicitation of  $w_1$  but before the elicitation of  $w_{2,1}$ . Such anchoring is consistent with prior results showing that valuations elicited using the BDM method can be affected by suggested prices or other anchors (Bohm, Lindén, and Sonnegård 1997; Mazar, Köszegi, and Ariely 2014). Thus, we do not believe this increase is relevant for a consumer welfare calculation, and we do not draw any substantive conclusion from it.

Figure 12 also illustrates  $\Delta w_2$ , the change in valuation of weeks 5–8 between midline and endline. The Control group's valuation increases, reflecting an unpredicted time effect. In open-answer questions, some people wrote that they were less willing to deactivate during the Thanksgiving holiday, and they may not have foreseen this as of the midline survey on October 11. By contrast, the Treatment group's valuation for weeks 5–8 decreases. Thus, the difference-in-differences  $\Delta w_2$  is negative.

We can estimate the difference-in-differences using the following regression:

$$(5) \quad \Delta w_{2,i} = \gamma D_i + \rho w_{1,i} + \nu_s + \varepsilon_i,$$

instrumenting for  $D_i$  with  $T_i$ . Table 6 presents results, winsorizing all WTAs at \$170 in column 1 and at \$1,000 in column 2. Relative to the Control group, the Treatment group reduced its post-endline valuation by \$14 to \$18, or about 14 percent of the Treatment group's average  $w_{2,1}$ . This suggests that deactivation eliminated projection bias or facilitated learning that reduced demand for Facebook by 14 percent. In turn, this suggests that the traditional estimates might somewhat overstate consumer surplus.

This result is consistent with our finding in Section IVD that deactivation reduced post-experiment Facebook use. However, because the WTA update  $\Delta w_2$  is unexpected, it suggests that the results from Section IVD may not be entirely explained by a “rational” habit formation model such as Becker and Murphy (1988), in which people foresee how consumption affects future marginal utility. Instead, these results suggest that at least some of the reduced Facebook demand caused by deactivation is driven by unexpected factors such as projection bias and learning.

One caveat is that the anchoring effect described above could affect our estimate of  $\gamma$ . If anchoring has the same effects on  $w_{2,1}$  and  $w_{2,2}$  in the Treatment group, then  $\Delta w_2$  is unaffected, and our estimate of  $\gamma$  is unbiased. If the anchoring effects decay between midline and endline, this would bias  $\hat{\gamma}$  away from zero, meaning that the true  $\gamma$  would be less than our estimate.<sup>36</sup> This would further strengthen our result that the valuation update caused by deactivation equals only a small share of valuations.

One interpretation of these results is that they reinforce the standard model calculation that Facebook generates many billions of dollars in consumer surplus. Another interpretation is that they further highlight why standard consumer surplus calculations based on elicited valuations can be problematic.

<sup>36</sup> An alternative experimental design choice we considered was to elicit  $w_{2,1}$  *before* revealing the weeks 1–4 offer price, separately for the case in which the participant would be paid to deactivate for weeks 1–4 and the case in which the participant would not be paid to deactivate. In this case, however, any anchoring effect would have appeared on  $w_{2,2}$  but not  $w_{2,1}$ , generating an unambiguous spurious treatment effect on  $\Delta w_2$ .

TABLE 6—CHANGE IN FACEBOOK VALUATION AFTER DEACTIVATION

	(1)	(2)
Share of time deactivated	−14.36 (2.60)	−18.22 (7.73)
Observations	1,634	1,634
Winsorized maximum WTA	170	1,000
Treatment mean weeks 5–8 WTA at midline	103	135

*Notes:* This table presents estimates of equation (5). The dependent variable is the change in WTA for post-endline deactivation measured at endline versus midline. Standard errors are in parentheses.

## VI. Conclusion

Our results leave little doubt that Facebook provides large benefits for its users. Even after a four-week “detox,” our participants spent substantial time on Facebook every day and needed to be paid large amounts of money to give up Facebook. Our results on news consumption and knowledge suggest that Facebook is an important source of news and information. Our participants’ answers in free response questions and follow-up interviews make clear the diverse ways in which Facebook can improve people’s lives, whether as a source of entertainment, a means to organize a charity or an activist group, or a vital social lifeline for those who are otherwise isolated. Any discussion of social media’s downsides should not obscure the basic fact that it fulfills deep and widespread needs.

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

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

## REFERENCES

- Acland, Dan, and Matthew Levy. 2015. “Naiveté, Projection Bias, and Habit Formation in Gym Attendance.” *Management Science* 61 (1): 146–60.
- Allcott, Hunt, and Matthew Gentzkow. 2017. “Social Media and Fake News in the 2016 Election.” *Journal of Economic Perspectives* 31 (2): 211–36.



- Alter, Adam. 2018. *Irresistible: The Rise of Addictive Technology and the Business of Keeping Us Hooked*. New York: Penguin Press.
- American National Election Studies. 2016. "2016 Time Series Study." <https://electionstudies.org/data-center/2016-time-series-study/>.
- Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–95.
- Appel, Helmut, Alexander L. Gerlach, and Jan Crusius. 2016. "The Interplay between Facebook Use, Social Comparison, Envy, and Depression." *Current Opinion in Psychology* 9: 44–49.
- Argyle, Michael. 2001. *The Psychology of Happiness*. London: Routledge.
- Athey, Susan, Julie Tibshirani, and Stefan Wagner. 2019. "Generalized Random Forests." *Annals of Statistics* 47 (2): 1148–78.
- Baker, David A., and Guillermo Perez Algorta. 2016. "The Relationship between Online Social Networking and Depression: A Systematic Review of Quantitative Studies." *Cyberpsychology, Behavior, and Social Networking* 19 (11).
- Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348 (6239): 1130–32.
- Bartels, Larry M. 1993. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87 (2): 267–85.
- Baym, Nancy K., Kelly B. Wagman, and Christopher J. Persaud. Forthcoming. "Mindfully Scrolling: Rethinking Facebook after Time Deactivated." *Social Media + Society*.
- Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 1991. "Rational Addiction and the Effect of Price on Consumption." *American Economic Review* 81 (2): 237–41.
- Becker, Gary S., and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4): 675–700.
- Becker, Gordon M., Morris H. Degroot, and Jacob Marschak. 1964. "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science* 9 (3): 226–32.
- Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones. 2012. "What Do You Think Would Make You Happier? What Do You Think You Would Choose?" *American Economic Review* 102 (5): 2083–2110.
- Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. "Adaptive Linear Step-Up Procedures that Control the False Discovery Rate." *Biometrika* 93 (3): 491–507.
- Besley, Timothy, and Robin Burgess. 2001. "Political Agency, Government Responsiveness and the Role of the Media." *European Economic Review* 45 (4–6): 629–40.
- Bohm, Peter, Johan Lindén, and Joakim Sonnegård. 1997. "Eliciting Reservation Prices: Becker-DeGroot-Marschak Mechanisms vs. Markets." *Economic Journal* 107 (443): 1079–89.
- Bolier, Linda, Merel Haverman, Gerben J. Westerhof, Heleen Riper, Filip Smit, and Ernst Bohlmeijer. 2013. "Positive Psychology Interventions: A Meta-Analysis of Randomized Controlled Studies." *BMC Public Health* 13: 119.
- Boxell, Levi. 2018. "Demographic Change and Political Polarization in the United States." <https://ssrn.com/abstract=3148805>.
- Brynjolfsson, Erik, Felix Eggers, and Avinash Gannamaneni. 2018. "Using Massive Online Choice Experiments to Measure Changes in Well-Being." NBER Working Paper 24514.
- Brynjolfsson, Erik, and Joo Hee Oh. 2012. "The Attention Economy: Measuring the Value of Free Digital Services on the Internet." Unpublished.
- Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2019. "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb, 23–60. Chicago: University of Chicago Press.
- Brynjolfsson, Erik, and Adam Saunders. 2009. "What the GDP Gets Wrong (Why Managers Should Care)." *MIT Sloan Management Review* 51 (1): 95.
- Burke, Moira, and Robert E. Kraut. 2014. "Growing Closer on Facebook: Changes in Tie Strength through Social Network Site Use." *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 4187–96.
- Burke, Moira, and Robert E. Kraut. 2016. "The Relationship between Facebook Use and Well-Being Depends on Communication Type and Tie Strength." *Journal of Computer-Mediated Communication* 21 (4): 265–81.
- Burke, Moira, Robert E. Kraut, and Cameron Marlow. 2011. "Social Capital on Facebook: Differentiating Uses and Users." *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 571–80.

- Burke, Moira, Cameron Marlow, and Thomas Lento. 2010. "Social Network Activity and Social Well-Being." *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 1909–12.
- Busse, Meghan R., Devin G. Pope, Jaren C. Pope, and Jorge Silva-Risso. 2015. "The Psychological Effect of Weather on Car Purchases." *Quarterly Journal of Economics* 130 (1): 371–414.
- Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. 2016. "Does the United States Have a Productivity Slowdown or a Measurement Problem?" *Brookings Papers on Economic Activity* (Spring): 109–82.
- Charness, Gary, and Uri Gneezy. 2009. "Incentives to Exercise." *Econometrica* 77 (3): 909–31.
- Chopik, William J. 2017. "Associations among Relational Values, Support, Health, and Well-Being across the Adult Lifespan." *Personal Relationships* 24 (2): 408–22.
- Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang. 2007. "Projection Bias in Catalog Orders." *American Economic Review* 97 (4): 1217–49.
- Corrigan, Jay R., Saleem Alhabash, Matthew Rousu, and Sean B. Cash. 2018. "How Much Is Social Media Worth? Estimating the Value of Facebook by Paying Users to Stop Using It." *PLOS ONE* 13 (12): e0207101.
- Csikszentmihalyi, Mihaly, and Reed Larson. 2014. "Validity and Reliability of the Experience-Sampling Method." In *Flow and the Foundations of Positive Psychology*, 35–54. New York: Springer.
- DellaVigna, Stefano, and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence." *Annual Review of Economics* 2: 643–69.
- DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics* 122 (3): 1187–1234.
- DellaVigna, Stefano, and Eliana La Ferrara. 2015. "Economic and Social Impacts of the Media." In *Handbook of Media Economics*, Vol. 1A, edited by Simon P. Anderson, David Strömberg, and Joel Waldfogel, 723–68. Amsterdam: Elsevier.
- Deters, Fenne Große, and Matthias R. Mehl. 2013. "Does Posting Facebook Status Updates Increase or Decrease Loneliness? An Online Social Networking Experiment." *Social Psychological and Personality Science* 4 (5): 579–86.
- Diener, Ed, Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. "The Satisfaction with Life Scale." *Journal of Personality Assessment* 49 (1): 71–75.
- Ellison, Nicole B., Charles Steinfield, and Cliff Lampe. 2007. "The Benefits of Facebook 'Friends': Social Capital and College Students' Use of Online Social Network Sites." *Journal of Computer-Mediated Communication* 12 (4): 1143–68.
- Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2018. "Social Media and Protest Participation: Evidence from Russia." <https://ssrn.com/abstract=2696236>.
- Enikolopov, Ruben, and Maria Petrova. 2015. "Media Capture: Empirical Evidence." In *Handbook of Media Economics*, Vol. 1A, edited by Simon P. Anderson, David Strömberg, and Joel Waldfogel, 687–700. Amsterdam: Elsevier.
- Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. 2011. "Media and Political Persuasion: Evidence from Russia." *American Economic Review* 101 (7): 3253–85.
- Facebook. 2016. "Facebook Q1 2016 Results: Earnings Call Transcript." *Seeking Alpha*, April 27, 2016. <https://seekingalpha.com/article/3968783-facebook-fb-mark-elliott-zuckerberg-q1-2016-results-earnings-call-transcript>.
- Facebook. 2018. "Facebook Reports Third Quarter 2018 Results." Press Release. <https://investor.fb.com/investor-news/press-release-details/2018/Facebook-Reports-Third-Quarter-2018-Results/default.aspx>.
- Fiorina, Morris P., and Samuel J. Abrams. 2008. "Political Polarization in the American Public." *Annual Review Political Science* 11: 563–88.
- Frison, Eline, and Steven Eggermont. 2015. "Toward an Integrated and Differential Approach to the Relationships between Loneliness, Different Types of Facebook Use, and Adolescents' Depressed Mood." *Communication Research*. doi.org/10.1177/0093650215617506.
- Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Economic Journal: Applied Economics* 8 (4): 160–88.
- Gentzkow, Matthew. 2006. "Television and Voter Turnout." *Quarterly Journal of Economics* 121 (3): 931–72.
- Gentzkow, Matthew. 2016. "Polarization in 2016." Unpublished.
- Gentzkow, Matthew, and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126 (4): 1799–1839.

- Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw. 2011. "How Large and Long-Lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *American Political Science Review* 105 (1): 135–50.
- Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (3): 653–63.
- Gerber, Alan S., Dean Karlan, and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1 (2): 35–52.
- Gerson, Jennifer, Anke C. Plagnol, and Philip J. Corr. 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." *Quarterly Journal of Economics* 116 (4): 1261–1303.
- Howard, Philip N., Aiden Duffy, Deen Freelon, Muzammil M. Hussain, Will Mari, and Marwa Maziad. 2011. "Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?" <https://ssrn.com/abstract=2595096>.
- Huber, Gregory A., and Kevin Arceneaux. 2007. "Identifying the Persuasive Effects of Presidential Advertising." *American Journal of Political Science* 51 (4): 957–77.
- Hughes, Mary Elizabeth, Linda J. Waite, Louise C. Hawkey, and John T. Cacioppo. 2004. "A Short Scale for Measuring Loneliness in Large Surveys: Results from Two Population-Based Studies." *Research on Aging* 26 (6): 655–72.
- 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 (3): 301–15.
- Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol. 2016. "Handwashing and Habit Formation." Unpublished.
- Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology: Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3): 405–31.
- Iyengar, Shanto, and Sean J. Westwood. 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3): 690–707.
- Kahneman, Daniel, Alan B. Krueger, David Schkade, Norbert Schwarz, and Arthur A. Stone. 2006. "Would You Be Happier If You Were Richer? A Focusing Illusion." *Science* 312 (5782): 1908–10.
- Kirkpatrick, David. 2011. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. New York: Simon & Schuster.
- Krasnova, Hanna, Helena Wenninger, Thomas Widjaja, and Peter Buxmann. 2013. "Envy on Facebook: A Hidden Threat to Users' Life Satisfaction." Unpublished.
- Kross, Ethan, Philippe Verduyn, Emre Demiralp, Jiyoung Park, David Seungjae Lee, Natalie Lin, Holly Shablock, John Jonides, and Oscar Ybarra. 2013. "Facebook Use Predicts Declines in Subjective Well-Being in Young Adults." *PLOS ONE* 8 (8): e69841.
- Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–1102.
- Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118 (4): 1209–48.
- Lyubomirsky, Sonja, and Heidi S. Lepper. 1999. "A Measure of Subjective Happiness: Preliminary Reliability and Construct Validation." *Social Indicators Research* 46 (2): 137–55.
- Martin, Gregory J., and Ali Yurukoglu. 2017. "Bias in Cable News: Persuasion and Polarization." *American Economic Review* 107 (9): 2565–99.
- Mazar, Nina, Botond Köszegi, and Dan Ariely. 2014. "True Context-Dependent Preferences? The Causes of Market-Dependent Valuations." *Journal of Behavioral Decision Making* 27 (3): 200–208.
- Molla, Rani, and Kurt Wagner. 2018. "People Spend Almost as Much Time on Instagram as They Do on Facebook." *Recode*, June 25, 2018. <https://www.recode.net/2018/6/25/17501224/instagram-facebook-snapchat-time-spent-growth-data>.
- Mosquera, Roberto, Mofioluwademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie. 2018. "The Economic Effects of Facebook." <https://ssrn.com/abstract=3312462>.
- Müller, Karsten, and Carlo Schwarz. 2018. "Fanning the Flames of Hate: Social Media and Hate Crime." <https://ssrn.com/abstract=3082972>.
- Myers, David G. 2000. "The Funds, Friends, and Faith of Happy People." *American Psychologist* 55 (1): 56–67.
- Nakamura, Leonard I., Jon Samuels, and Rachel H. Soloveichik. 2016. "Valuing 'Free' Media in GDP: An Experimental Approach." Federal Reserve Bank of Philadelphia Working Paper 16-24.

- Napoli, Philip M. 2014. "Measuring Media Impact." The Norman Lear Center. <https://learcenter.org/pdf/measuringmedia.pdf>.
- Newport, Cal. 2019. *Digital Minimalism: Choosing a Focused Life in a Noisy World*. New York: Penguin.
- Olken, Benjamin A. 2009. "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages." *American Economic Journal: Applied Economics* 1 (4): 1–33.
- Pew Research Center. 2018. "News Use across Social Media Platforms 2018." <http://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/>.
- Reis, Harry T., W. Andrew Collins, and Ellen Berscheid. 2000. "The Relationship Context of Human Behavior and Development." *Psychological Bulletin* 126 (6): 844–72.
- Sagioglu, Christina, and Tobias Greitemeyer. 2014. "Facebook's Emotional Consequences: Why Facebook Causes a Decrease in Mood and Why People Still Use It." *Computers in Human Behavior* 35: 359–63.
- Satici, Seydi Ahmet, and Recep Uysal. 2015. "Well-Being and Problematic Facebook Use." *Computers in Human Behavior* 49: 185–90.
- Settle, Jaime E. 2018. *Frenemies: How Social Media Polarizes America*. New York: Cambridge University Press.
- Shakya, Holly B., and Nicholas A. Christakis. 2017. "Association of Facebook Use with Compromised Well-Being: A Longitudinal Study." *American Journal of Epidemiology* 185 (3): 203–11.
- Simonsohn, Uri. 2010. "Weather to Go to College." *Economic Journal* 120 (543): 270–80.
- Spenkuch, Jörg L., and David Toniatti. 2016. "Political Advertising and Election Outcomes." CESifo Working Paper 5780.
- Stone, Arthur A., and Saul Shiffman. 1994. "Ecological Momentary Assessment (EMA) in Behavioral Medicine." *Annals of Behavioral Medicine* 16 (3): 199–202.
- Strömberg, David. 2015. "Media and Politics." *Annual Review of Economics* 7: 173–205.
- Sunstein, Cass R. 2001. *Republic.com*. Princeton, NJ: Princeton University Press.
- Sunstein, Cass R. 2017. *#Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press.
- Sunstein, Cass R. Forthcoming. "Valuing Facebook." *Behavioural Public Policy*.
- Syverson, Chad. 2017. "Challenges to Mismeasurement Explanations for the US Productivity Slowdown." *Journal of Economic Perspectives* 31 (2): 165–86.
- Tandoc, Edson C., Patrick Ferrucci, and Margaret Duffy. 2015. "Facebook Use, Envy, and Depression among College Students: Is Facebooking Depressing?" *Computers in Human Behavior* 43: 139–46.
- Twenge, Jean M. 2017. *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy—and Completely Unprepared for Adulthood—and What That Means for the Rest of Us*. New York: Simon & Schuster.
- Twenge, Jean M., Thomas E. Joiner, Megan L. Rogers, and Gabrielle N. Martin. 2018. "Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates among U.S. Adolescents after 2010 and Links to Increased New Media Screen Time." *Clinical Psychological Science* 6 (1): 3–17.
- Twenge, Jean M., Gabrielle N. Martin, and W. Keith Campbell. 2018. "Decreases in Psychological Well-Being among American Adolescents after 2012 and Links to Screen Time during the Rise of Smartphone Technology." *Emotion* 18 (6): 765–80.
- Twenge, Jean M., and Heejung Park. 2019. "The Decline in Adult Activities among U.S. Adolescents, 1976–2016." *Child Development* 90 (2): 638–54.
- Twenge, Jean M., Ryne A. Sherman, and Sonja Lyubomirsky. 2016. "More Happiness for Young People and Less for Mature Adults: Time Period Differences in Subjective Well-Being in the United States, 1972–2014." *Social Psychological and Personality Science* 7 (2): 131–41.
- US Census Bureau. 2017. "2017 American Community Survey 1-Year Estimates." <https://www.census.gov/newsroom/press-kits/2018/acs-1year.html>.
- Vanden Abeele, Mariek M. P., Marjolijn L. Antheunis, Monique M. H. Pollmann, Alexander P. Schouten, Christine C. Liebrecht, Per J. van der Wijst, Marije A. A. van Amelsvoort et al. 2018. "Does Facebook Use Predict College Students' Social Capital? A Replication of Ellison, Steinfield, and Lampe's (2007) Study Using the Original and More Recent Measures of Facebook Use and Social Capital." *Communication Studies* 69 (3): 272–82.
- Verduyn, Philippe, David Seungjae Lee, Jiyoung Park, Holly Shablack, Ariana Orvell, Joseph Bayer, Oscar Ybarra, John Jonides, and Ethan Kross. 2015. "Passive Facebook Usage Undermines Affective Well-Being: Experimental and Longitudinal Evidence." *Journal of Experimental Psychology* 144 (2): 480–88.

**This article has been cited by:**

1. Luca Braghieri, Ro'ee Levy, Alexey Makarin. 2022. Social Media and Mental Health. *American Economic Review* **112**:11, 3660-3693. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
2. Sergei Guriev, Elias Papaioannou. 2022. The Political Economy of Populism. *Journal of Economic Literature* **60**:3, 753-832. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
3. Hunt Allcott, Matthew Gentzkow, Lena Song. 2022. Digital Addiction. *American Economic Review* **112**:7, 2424-2463. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]
4. Ro'ee Levy. 2021. Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *American Economic Review* **111**:3, 831-870. [[Abstract](#)] [[View PDF article](#)] [[PDF with links](#)]