

The Effects of Facebook Access during the 2022 French Presidential Election: Can We Incentivize Citizens to be Better Informed and Less Polarized?

Kevin Arceneaux¹, Martial Foucault¹, Kalli Giannelos^{1,3}, Jonathan Ladd², Can Zengin^{1,4}

¹Center for Political Research (CEVIPOF), Sciences Po, France

²McCourt School of Public Policy and Department of Political Science, Georgetown University, USA

³Paris-Dauphine University, France

⁴Temple University, USA

Abstract

Despite its early promise to empower and inform, many political and social observers fret that social media platforms are rife with polarizing content, misinformation, disinformation, and are designed to ensnare its users, threatening its users' mental health and the health of democracy. Randomized field experiments that incentivize users to deactivate their Facebook accounts offers credible evidence that access to Facebook increases political knowledge, reduces subjective well-being, and may either increase or decrease political polarization. We replicate and extend this research. In addition, incentivizing participants to deactivate their Facebook account, we also provided some participants with minimalist informational messages warning them about misinformation while providing digital literacy tips. A parallel line of lab and survey experiments shows that these types of minimalistic treatments can induce people to be more accurate and civil, and we investigate whether these treatments when paired with Facebook deactivation helps participants become more informed and less polarized. We conducted a preregistered randomized Facebook deactivation field experiment during the 2022 French presidential election. In line with previous research, we find that Facebook informs people about politics, reduces subjective well-being, but we find little evidence that it affects either social or political polarization. In contrast to previous research, we find that minimalistic informational treatments have limited effects on political knowledge and practically no effects on subjective well-being and polarization.

We thank the McCourt Institute at Sciences Po for its generous funding for this research and Céline Laffineur for her research assistance. We also thank Lene Aarøe, Michael Bailey, Andy Guess, Lala Muradova, Josh Tucker, Tiago Ventura, participants in colloquiums at Aarhus University, the Center for Political Research at Sciences Po (CEVIPOF), Gothenburg University, the McCourt Institute at Sciences Po, and the Universitat Pompeu Fabra. A version of this paper was presented at the 2023 Midwestern Political Science Association. All errors remain our own.

In the space of the past two decades, optimism that the Internet would provide a deliberative space for the free exchange of ideas and give voice and a tool for political organization to the less powerful and oppressed has given way to deep pessimism about the deleterious social and political effects of social media (Zhao 2014). Instead of reasoned, respectful deliberation about politics, messages on social media often include invective along with arguments put forth in bad faith that engender outrage and polarization (Bail 2022; Kubin and von Sikorski 2021; Rathje, Van Bavel, and Linden 2021; Settle 2018; Van Bavel, et al 2021). A central problem is that people are more likely to be exposed to information about news and politics that they may not have sought out on social media sites because all the most popular of these platforms, including Facebook, use algorithms to provide its user with diverse set of options that attract engagement (Bashky et al. 2015; Eady, et al. 2019; Fletcher, Robertson, and Nielsen 2021). Exposure to counter-attitudinal social media posts that come from acquaintances or friends of friends can increase polarization (Anspach 2016; Bail 2022; Rathje, Van Bavel, and Linden 2021 but see Levy 2021). Another related problem is that people with a strong interest in politics are more likely to use social media to seek out information about politics (Cacciatore et al. 2018), and those who do use social media to learn about politics often become more politically engaged in the process (Lee, Shin, and Hong 2018). On the one hand, those who desire to learn about politics, and are already rather informed and involved, might be led by stories they are fed by algorithms to become more engaged and polarized. On the other hand, most people do not always read information carefully while scrolling, diminishing the quality of some social media users' political knowledge (Anspach, Jennings, and Arceneaux 2019).

Much of the research on the effects of social media either come from observational studies that report correlations between social media use and political attitudes and behavior or from survey and lab experiments that isolate the effects of specific pieces of information in a stylized environment. Observational studies can provide an accurate description of how people behave on

social media, but it is difficult to make strong causal inferences from these kinds of studies given the degree of selectivity in social media settings. Survey experiments allow researchers to isolate the causal effects of encountering content in a stylized setting, but without a more complete assessment of the social media environments people encounter in the real world, it is difficult to know whether the internally valid causal effects observed in these experiments would manifest in real-world settings.

Randomized field experiments offer a way to address both limitations, because they use random assignment to isolate causal effects in real-world settings. Two innovative field experiments have employed a deprivation design that incentivized some participants to deactivate their Facebook accounts. The first was conducted during the 2018 midterm elections in the United States. These scholars found that giving up Facebook for a month increased happiness, reduced polarization, but also reduced levels of political engagement and knowledge (Allcott et al. 2020). A second, similar experiment, conducted outside of an electoral context in Bosnia Herzegovina during the summer of 2019, also found that those who gave up Facebook for a week were happier and less informed, but (unexpectedly) *more* polarized toward their ethnic outgroup (Asimovic et al. 2021).

Taken together, these findings suggest that people do learn about politics from Facebook, but at the potential cost of their own subjective well-being. At the same time, they also leave open to question whether Facebook polarizes. It may have a polarizing effect in the US (Allcott et al. 2020; Anspach 2016; Bail 2022; Settle 2018), but perhaps not in other contexts. Adding to the uncertainty, the two field experiments in this area looked at different countries and different types of polarization. Allcott et al. studied specifically the effects of Facebook on political polarization (partisan and ideological), while Asimovic et al. studied the effects of Facebook on social polarization (ethnicity). One plausible, albeit post-hoc, explanation for this pattern of results is that in all contexts Facebook increases political polarization, while reducing social polarization. Another

possibility is that cultural, political, or social differences between the US and Bosnia Herzegovina explain the difference. Only additional research can provide insight into the differences in findings between these first two studies.

Moreover, the deprivation experimental design used in these two previous studies simulates what happens when people are left to their own devices once they choose to give up Facebook while it continues to exist, as opposed to a world in which Facebook (or other social media) never existed. In the Allcott et al. experiment, participants in the deprivation treatment group reported spending more time with friends, alone watching television, and less time learning about the news than they would have otherwise. Many of these individuals were habituated to getting their news from Facebook and were not prepared to substitute it with other news sources, and the experience of giving up Facebook for a month left many participants wrestling with the desire to find a way to “mindfully” engage with the media environment (Baym et al. 2020).

A parallel stream of research, largely conducted in more controlled experimental settings, suggests that targeted interventions could help people use social media in a healthier manner. Providing people with advice on how to better use digital media can motivate them to reach accurate conclusions as opposed to ones that favor their favored positions or groups. These informational treatments, in turn, reduce levels of misinformation and polarization (e.g., Groenendyk and Krupnikov 2022; Pennycook et al. 2020; Ruggeri et al. 2022). A striking feature about this stream of research is these informational treatments are minimalist. Participants in these studies are merely requested to consider the accuracy of information or to observe norms regarding civility. If these types of treatments are effective it suggests that, perhaps, people *want* to use social media in ways that are more aligned with the utopian Internet envisioned by cyberoptimists, but they need to be given a nudge to do so.

A Direct Replication and an Extension

We conducted a similar randomized field experiment using a deprivation design, in which we asked a diverse sample of French Facebook users to deactivate their Facebook accounts during the 2022 French presidential election campaign. There were nearly 46 million adult Facebook users in France in 2022 (nearly 70% of its population).¹ France elects its president through a two-round system. The first-round election was held on 10 April, featuring 12 candidates. Since none of the candidates received a majority of the votes, the top two vote recipients — incumbent President Emmanuel Macron (centrist) and Marine Le Pen (far right) — competed in the second-round election held on 24 April. President Macron won that election with 58.5% of the validly casted votes.

Our experiment offers two contributions. First, we built into our experiment a direct replication of the Allcott et al. (2020) and Asimovic et al. (2021) experiments. Participants in our study were randomly assigned to a treatment arm in which they were compensated to deactivate Facebook for 25 days during the presidential election. Direct replications do more than verify the empirical scope of previous findings, they also facilitate theoretical refinement by confirming or disconfirming the scope of previous theoretical claims (Chambers 2017). France is similar and different in many ways from Bosnia Herzegovina and the US.² If we replicate the key findings of these studies, it offers further evidence that these effects of Facebook are robust across contexts.

¹ [https://napoleoncat.com/stats/facebook-users-in-france/2022/02/#:~:text=There%20were%2047%20495%20600,group%20\(12%20300%20000\).](https://napoleoncat.com/stats/facebook-users-in-france/2022/02/#:~:text=There%20were%2047%20495%20600,group%20(12%20300%20000).)

² Like the US, France is a western democracy with a presidential system, and similar to Bosnia Herzegovina the major political cleavage in France is ethno-religious (Northern African immigrants and their descendants who tend to identify as Muslims). Yet, there are important differences as well. France, unlike the US, has a multi-party system and political polarization does not reflect two opposing, equally sized party coalitions, and unlike Bosnia Herzegovina where Muslims and Christians each make up roughly 50% of the population, Christians and non-religious people in France far outnumber the adherents of Islam. In this way, France offers a less extreme case with respect to social and political polarization, while having important political cleavages nonetheless.

The results can also shed light on the reasons for discrepancies between previous studies, such as, in this case, the apparently different findings on polarization.

Second, we extended previous research by testing whether deactivating Facebook coupled with minimalist treatments aimed at helping people better navigate and process information that they encounter in both the online and offline world would counteract the negative effects of Facebook deactivation on political engagement and polarization found in previous studies. One could think of our approach as akin to studying the effects of a diet regimen coupled with tips for healthy eating. We designed informational treatments that built on the kinds of treatments featured in other studies. Half of the participants assigned to the deactivation treatment group were randomly assigned to receive four brief messages over the course of the study that informed them about: 1) the need to see the potentially addictive nature of social media, 2) the need to recognize the social (and thus public) nature of one's behavior on social media, 3) the need to seek out a diversity of information and consider the accuracy of the news that they find in order to counter attempts at spreading misinformation (a treatment similar to Pennycook et al. 2020), and 4) the need to foster a space with civility and tolerance. We preregistered our hypotheses regarding both treatments and how these treatments might interact with subjects' predisposition prior to receiving the data (<https://osf.io/xt5zg/>).

Methods

Participants

We engaged a well-respected international survey research firm to recruit French citizens who were eligible to vote and who reported that they regularly used Facebook, and responded to an invitation by saying that they would be willing to participate in our study by deactivating their Facebook account during for several weeks in exchange for 80€. Among those who said they would be willing

to do so, we asked if they would be willing to be entered into a lottery where they had a 50% chance of being in the deactivation study that involved completing two additional surveys (for which they would receive 80€) or a study where they would just take the two additional surveys in return for the regular incentive that the survey firm pays them for participating in surveys. Of the 2,246 respondents who said they would be willing to participate in such a lottery, we randomly assigned 1,117 to be in the deactivation treatment arm. We requested that they deactivate their Facebook accounts in exchange for the agreed upon 80€ payment for a little more than three weeks, starting 10 days prior to the first-round election (10 April 2022) and lasting until the day after the second-round election (25 April 2022). Following the protocol of previous deactivation studies, we engaged the survey research firm to check daily the Facebook accounts of participants in the treatment arm to verify that their accounts were indeed deactivated and to send a message reminding participants of their commitment to deactivate their Facebook account if they had not done so. Participants were surveyed at three times during the study: before the study began (1 April), in between the first- and second-round elections (19 April, what we will call the “midline” survey), and after the second-round election (29 April). Participants were paid for their participation after completing all three survey waves. See the Supplemental Information (SI) for descriptive statistics.

Procedures

After participants completed the baseline survey, they were informed whether they had been randomly assigned to the deactivation treatment or the control group. Those assigned to the treatment group were instructed how to deactivate their Facebook account and we explained that deactivation did not delete their data (they could reactivate after the study) and that they would continue to have access to Facebook Messenger. The research firm tasked employees with pinging the Facebook accounts of those assigned to the treatment to ensure that the accounts were

deactivated. If an account had been reactivated, the research firm sent a message to the participant reminding them that they had agreed to deactivate their account for 80€ and requesting that they comply with the protocol.

Of the 1,117 participants randomly assigned to the deactivation treatment arm, 547 were assigned to the pure deactivation treatment, which replicates previous studies, and 570 were assigned both to deactivate their accounts and to receive four informational messages during the experiment. The messages were delivered via email on fixed days. The addiction message was sent on 12 April 2022, the privacy message was sent on 14 April (both before the midline survey), the accuracy message was sent on 21 April, and the civility message was sent on 22 April (before the second-round election). See the SI for the full wording of the messages.

Outcome Measures

Subjective Well-being. Participants were asked on the midline and endline surveys several questions that tapped their subjective sense of well-being. We asked people to evaluate “how satisfied are you in the life that you lead” on an 11-point scale ranging from 0 (labeled “not at all satisfied”) to 10 (labeled “absolutely satisfied”) and 5 labeled “neither satisfied nor dissatisfied.” In addition to making this general evaluation, participants were asked to self-report how much they felt several positive and negative discrete emotions on a 11-point scale ranging from 0 to 10 (“How much have you felt these emotions over the past two weeks?”). The list of emotions included: joy, fulfillment, anxiety, boredom, loneliness, depression, and isolation. We created a well-being index by taking the average of participants’ responses to the life satisfaction question and seven discrete emotion questions (the negative emotion items — anxiety, boredom, loneliness, depression, and isolation — were reverse coded so that larger numbers indicated positive well-being). All of these outcomes were

measured by Asimovic et al., whereas Allcott et al. measured happiness (similar to joy), loneliness, depression, anxiety, boredom, and life satisfaction.

Political Engagement. We included two measures of political engagement on the midline and endline surveys. One measured online political engagement simply: “Have you recently engaged in online discussions on social media about the elections?” (yes/no). The other measured how much time people spent recently following the news: “Last week, how many minutes did you spend reading/watching/listening to news about politics, including news on social media?” The response set offered participants the following options: 0 minutes, less than 30 minutes, between 30 and 60 minutes, between 1 and 2 hours, and 2 hours or more (similar to Allcott et al’s measure).

News Knowledge. Following both Asimovic et al and Allcott et al’s approach, we asked participants on the midline and endline surveys to read 12 headlines and tell us whether each was true or false. Of the 12 headlines, six described events that had actually been reported in the news in the past week, while the other six were written in the style of a fake news headline: plausible, but false. In addition, we chose or created six headlines that were about politics and six that were about entertainment or sports. Thus, we ended up with three true statements about politics, three false statements about politics, three true statements about entertainment/sports, and three false statements about entertainment/sports.

Affective Polarization. On the midline and endline surveys, participants were asked to rate how they felt about members of different groups on an 11-point scale ranging from 0 (“you don’t like them at all”) to 10 (“You have a very warm feeling toward members of this group”). For the purpose of this project, we measured three forms of affective polarization: 1) partisan, 2) ideological, and 3) social.

Similar to Allcott et al., we measured partisan and ideological polarization by taking the difference between the rating of a participant's ingroup (their preferred party or ideological group) and outgroup (their rating of their least preferred party or opposing ideological group). Similar to Asimovic, et al, we measured social polarization by taking the difference between participants' feelings toward citizens without an immigrant background as well as their feelings toward citizens with a Maghreb and African immigration background, which represents an important social (cultural and religious) cleavage in France.

See the SI for all complete descriptive statistics and question wordings in French, and for the wordings of the headlines that we showed participants in the midline and endline surveys (the headlines are different, since they reflect current events from the previous week).

Manipulation Check

As a manipulation check, Figure 1 shows the effect of treatment assignment on self-reported Facebook use on the midline and endline surveys. Participants were asked "How often do you currently check your Facebook feed?" and the response set included: 0 times per week, once per week, once per day, more than once per day, and more than 10 times a day. The treatments reduced Facebook use considerably in the midline wave (treatment – control = -1.1 for both treatment arms, $p < 0.01$) and endline waves (Deactivation Only – Control = 0.78; Deactivation + Information – Control = 0.73, $p < 0.01$). On the midline survey, 37.7% of the Deactivation Only group and 35% of the Deactivation + Information group reported not checking their Facebook feed at all as required by the experimental protocol, compared to just 0.01% of the control group. Given that the survey firm checked that the treatment group participants' Facebook accounts were deactivated, we are not sure why many other continued reporting Facebook use. One possibility is that they interpreted the question as asking about their general behavior in the past, including before the

experiment started. Another possibility is that some participants found a way around the experimental protocol. Whatever the case, noncompliance does not vitiate the internal validity of our experimental design.

Almost all field experiments contend with the issue of non-compliance (Gerber and Green 2012). In many ways this is a feature and not a bug of these experiments as it allows researchers to evaluate the effects of real-world treatments where people have agency over their behavior. Most importantly, non-compliance does not affect the internal validity of our experimental design, since we report intent-to-treat effects that compare the behavior of randomly assigned group members, irrespective of their compliance with the experimental protocol.

In endline survey, a partial explanation for why some participants in the treatment group reported using Facebook is that they were allowed to do so, since the experiment ended five days before the survey was conducted. In the endline survey, 20% of treatment group participants reported that they had opted to continue to forgo using Facebook. Of those who reported on the midline survey that they were not currently checking their Facebook feed, 33.3% in the Deactivation Only group and 35.3% in the Deactivation + Information group reported returning to pre-study levels of Facebook use. Of those, who reported that they quit using Facebook during the study (i.e., they fully complied with the treatment), 58.9% of the Deactivation Only group and 61.2% of Deactivation + Information group reported returning to Facebook but at lower levels than before the study began.

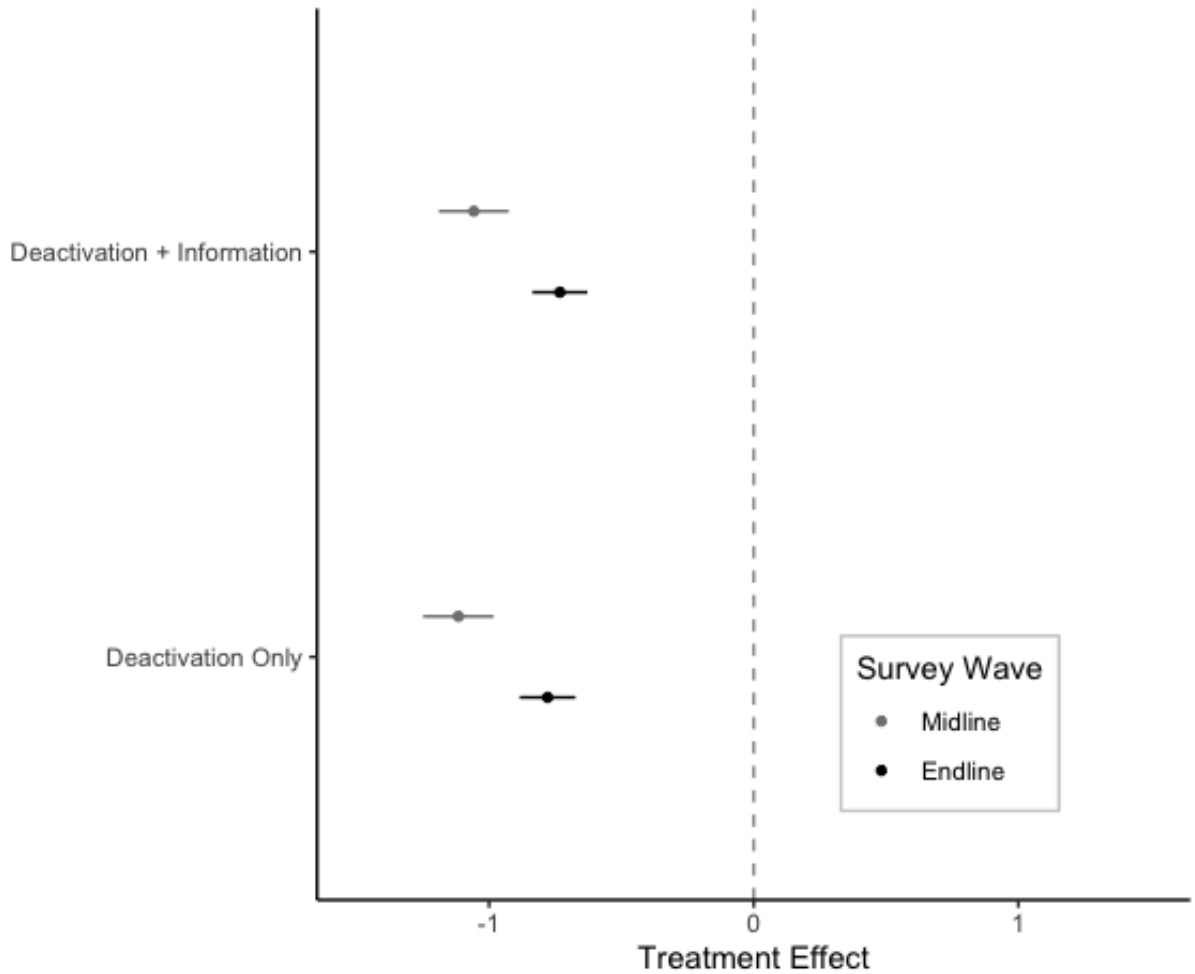


Figure 1: Manipulation Check for Self-reported Time Spent on Facebook

Note: The dots indicate the difference between the treatment and control group and horizontal lines represent 95% confidence intervals.

With respect to the informational treatments, we asked participants randomly assigned to the deactivation + information group if they “received informative messages by email indicating the negative effects of social media and online platforms?” Of the 570 participants, 74% reported that they had, 72.6% reported that they read them, and 68.6% reported that they followed their suggested guidelines.

Preregistered Hypotheses and Analysis Plan

For this project, we pre-registered several findings in line with the previous literature:

H1a. Participants in the Deactivation Only treatment condition will report higher levels of subjective well-being relative to participants in the control group.

H1b. Participants in the Deactivation + Information treatment condition will report higher levels of subjective well-being relative to participants in the control group and those in the Deactivation Only treatment condition.

H2a. Participants in the Deactivation Only treatment condition will report lower levels of news knowledge relative to participants in the control group.

H2b. Participants in the Deactivation + Information treatment condition will report higher levels of new knowledge relative to participants in the Deactivation Only treatment condition.

H3a. Participants in the Deactivation Only treatment condition will report lower levels of affective polarization relative to participants in the control group.

H3b. Participants in the Deactivation + Information treatment condition will report lower levels of affective polarization relative to participants in the Deactivation Only treatment condition.

For each outcome variable, we preregistered the following OLS regression model for testing the hypotheses above:

$$y_i = \beta_0 + \beta_1 DO_i + \beta_2 DI_i + \beta_3 y_{it=1} + e_i,$$

where y_i = the outcome variable for each participant i , DO = an indicator variable $\{0, 1\}$ for those assigned to the Deactivation Only treatment condition, DI = an indicator variable $\{0, 1\}$ for those assigned to the Deactivation + Information treatment condition, $y_{it=1}$ = the outcome variable for each participant recorded in the baseline wave (if available), and e = error term. For ease of interpretation, we report coefficient plots in the main text, but the regression models for all the

analyses the follow can be found in the SI. As noted above, we also test several hypotheses about heterogeneous treatment effects (i.e., how our treatment effects interact with predispositions), which are listed in our pre-registration report. (<https://osf.io/xt5zg/>).

Findings

Substitution Behavior

We asked participants in the treatment arm on the midline survey how they spent most of their time and we gave them several options from which to choose. Figure 2 reports the proportion of participants who fell into each category. For the most part, participants in the treatment arm reported spending more time with friends and family as well as watching television. There do not appear to be major differences between the Deactivation Only and the Deactivation + Information conditions. Both groups were just as likely to consume other social media (Twitter and Instagram, specifically), spent time on the Internet, read a newspaper, read news online, or to not use technology. While it appears that the Deactivation + information group was more likely to watch television than the Deactivation Only group (41.7% versus 36.4%) and the Deactivation Only group was more likely to spend time with their friends and family than the Deactivation + Information group (31.2% versus 26.5%), neither of these differences are statistically significant, which means that we cannot rule out sampling variability as a possible explanation.

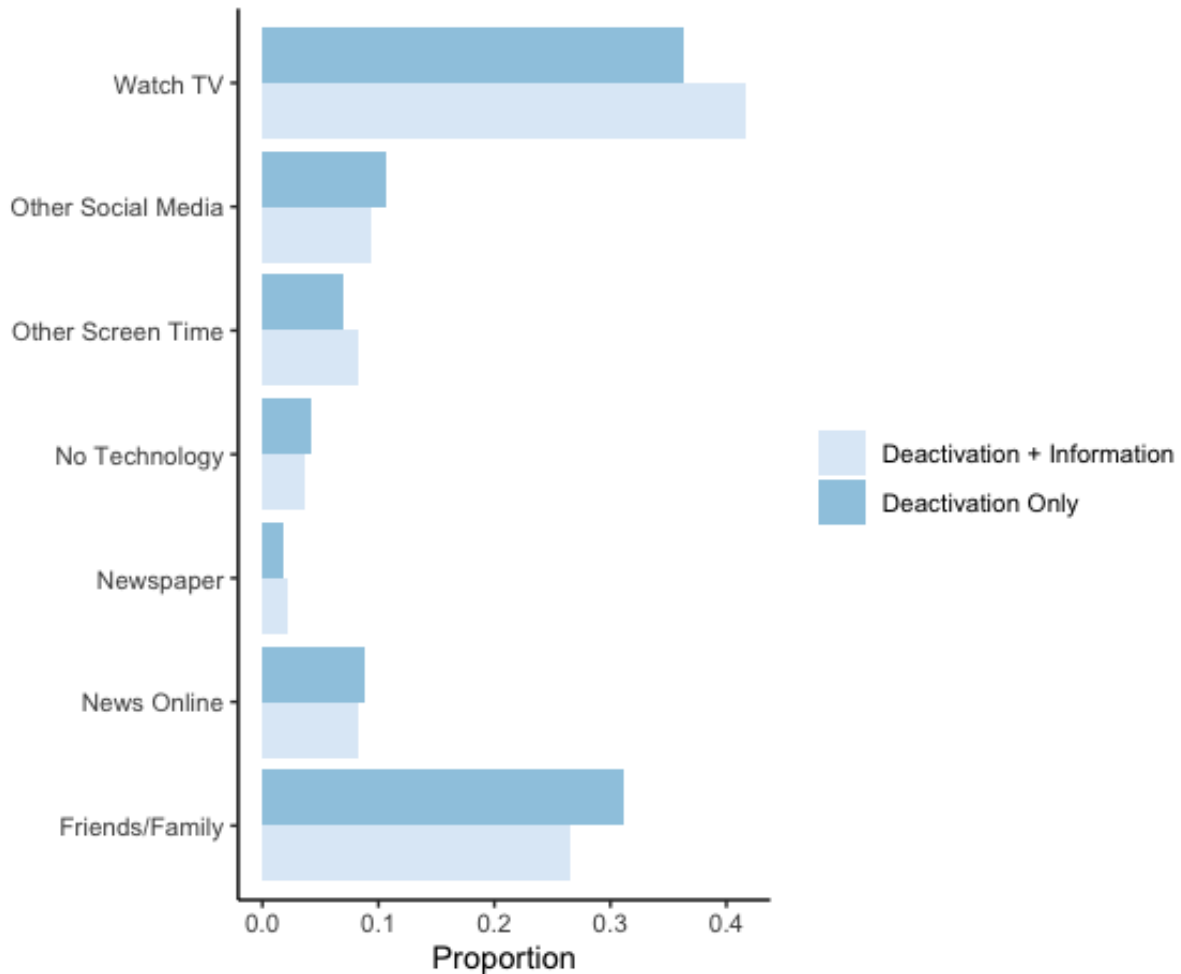


Figure 2: Where Participants in the Treatment Arms Reported Spending Most of their Time on the Midline Survey

Subjective Well-being

Figure 3 summarizes the effects Facebook deactivation on our measures of subjective well-being. The results are largely in line with previous research. By the end of the study, those in the Deactivation Only group rated their overall well-being, measured by the well-being index, more positively than those in the control group ($d = 0.05$, $p < 0.05$), which mirrors Allcott et al. and Asimovic et al's findings. Those who deactivated Facebook for nearly a month reported feeling

more joy, fulfilled, better life satisfaction, less anxiety, less boredom, less loneliness, less isolation, and less depression by the end of the study.

Next, we turn to the midline survey, to see what effects were present after only 19 days and before the campaign was completed. The positive effects of the Deactivation only treatment were evident on the midline survey with respect to the positive indicators (joy, fulfilled, and life satisfaction), but less so with respect to the negative indicators (anxiety, boredom, loneliness, isolation, and depression). In fact, the immediate effect of deactivating Facebook was to increase anxiety and boredom, even though this effect reversed itself by the end of the campaign.

Looking at the entire one-month campaign period, French citizens who were paid to deactivate Facebook for felt better at the end of the study, just as Americans and Bosnia Herzegovinians who deactivated their Facebook accounts. These findings are consistent with H1a and corroborate previous research.

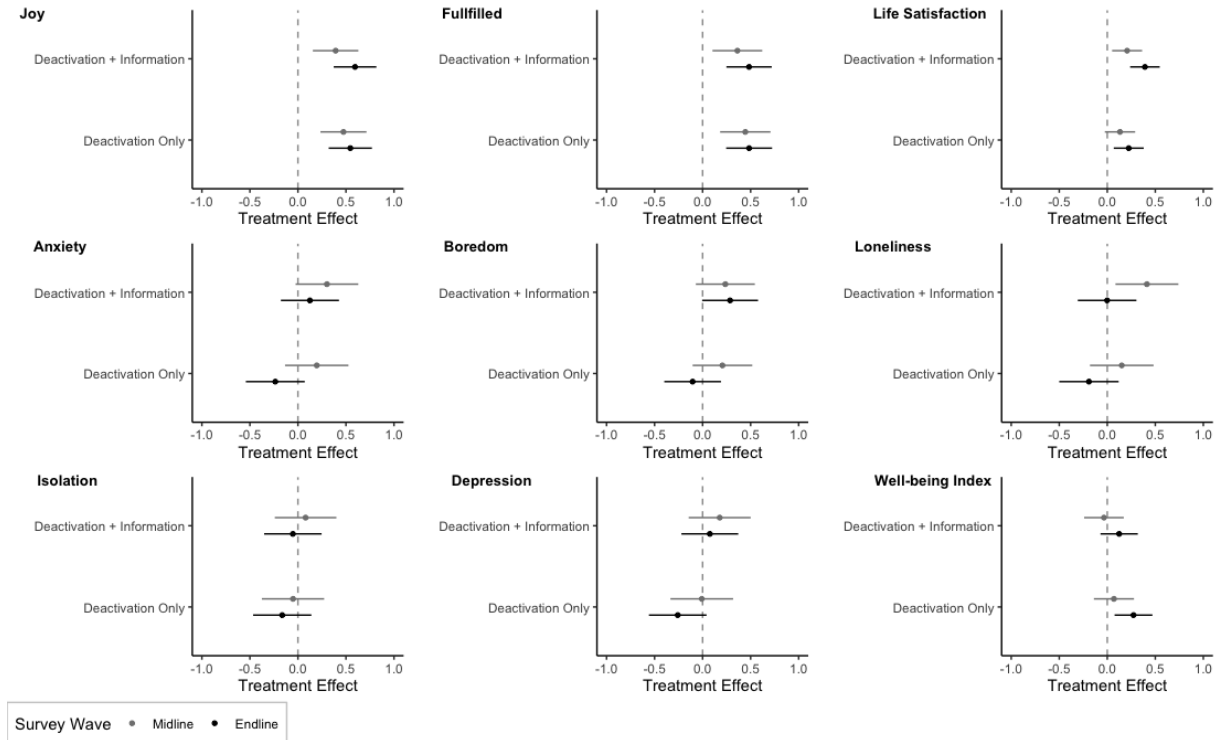


Figure 3: The Effects of Facebook Deactivation on Subjective Well-being

Note: The dots indicate the difference between the treatment and control group and horizontal lines represent 95% confidence intervals.

In contrast, we do not find consistent evidence for H1b. While the Deactivation + Information group also reported higher levels of well-being by the end of the study relative to the control group ($p < 0.05$), we do not observe that they report higher levels of well-being than those in the Deactivation Only group. If anything, by the endline survey, those in the Deactivation + Information reported felling higher levels of anxiety and boredom than those in the Deactivation Only group.

News Knowledge and Political Engagement

In line with Alcott et al. and Asimovic et al., Facebook deactivation decreased participants knowledge about *political* news ($d = -0.07$, $p < 0.05$) on the endline survey (Figure 4). Interestingly, it

had no effects on people’s knowledge about sports and entertainment events (non-political news). These results are consistent with H2a and replicate previous research. H2b anticipated that the information treatments would counteract the negative effects of Facebook deactivation on news knowledge. The results partially corroborate this prediction. On the midline survey, those who received the informational messages in the deactivation treatment condition, knew *more* about politics relative to the Deactivation Only group *and* the control group ($d = 0.08, p < 0.05$). However, by the endline survey, those in the Deactivation + Information group knew *less* about political news than the control group, similar to the Deactivation Only group.

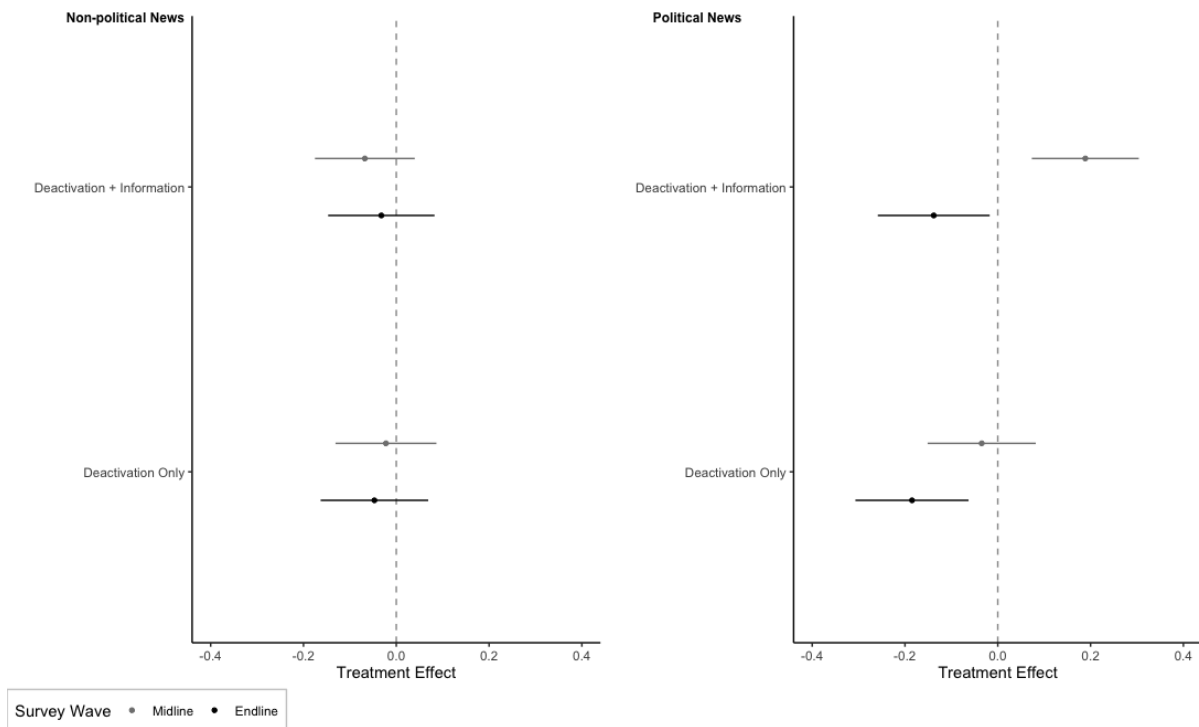


Figure 4: The Effects of Facebook Deactivation on News Knowledge

Note: The dots indicate the difference between the treatment and control group and horizontal lines represent 95% confidence intervals.

When we dig deeper, we find that the negative effects of the Deactivation Only treatment on news knowledge on the endline survey were largely driven by participants' diminished ability to identify true political news ($p < 0.05$). Facebook deactivation had essentially no effect on people's belief in false news. Put differently, participants in the control group – who continued to use Facebook – were *not* more likely to believe fake news headlines relative to those who deactivated Facebook. Likewise, the Deactivation + Informational treatment increased participants' knowledge of true political news in the midline survey, yet this positive effect dissipated by the endline survey. By the end of the study, informational treatments had essentially no additional effect on people's ability to identify fake news.

A possible explanation for the negative effects of Facebook deactivation on political news knowledge is that those in the deactivation treatment arm reported spending less time following the news in the previous week ($p < 0.05$, see the SI). Participants in the Deactivation + Information treatment condition were also less likely to follow the news in the previous week, which indicates that the informational treatments did not nudge people who deactivated Facebook to seek it out through other means (other social media, offline news, etc.).

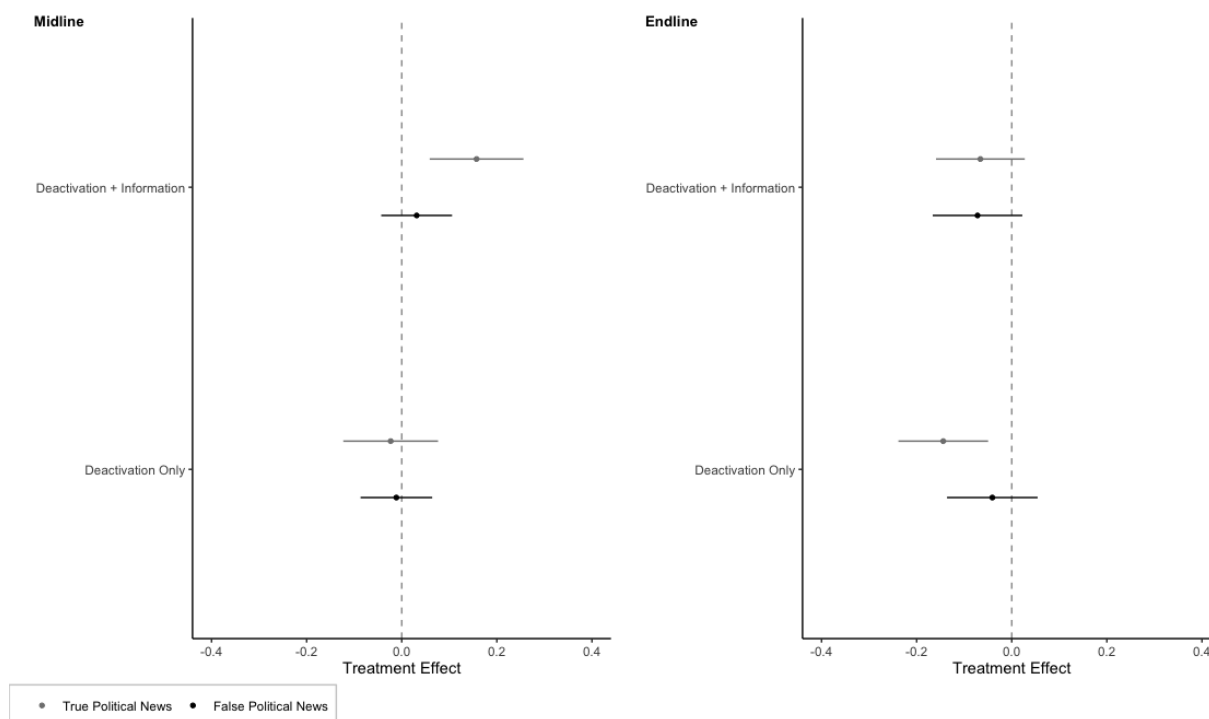


Figure 5: The Effects of Facebook Deactivation on Discernment Between True and False Political News

Note: The dots indicate the difference between the treatment and control group and horizontal lines represent 95% confidence intervals.

Affective Polarization

Recall that H3a and H3b predicted that Facebook deactivation would reduce affective polarization, which is in line with Alcott et al., who found that Facebook deactivation reduced affective political polarization, and in contrast with Asimovic et al., who found (contrary to their preregistered expectations) that Facebook deactivation increased social polarization. Our results, shown in Figure 6, contradict both studies. We find that Facebook deactivation had essentially no effect on partisan polarization, ideological polarization, or social polarization (feelings toward those with and without Maghreb or African ancestry). Not only are all of the effects that we observed statistically insignificant ($p > 0.05$), the effect size in almost every case is near zero. The only exception to this

pattern is that the Deactivation + Information treatment may have slightly increased partisan polarization on the midline survey – although this effect is not statistically significant and entirely dissipates by the end of the campaign. In sum, we find no evidence that either Facebook deactivation or Facebook deactivation supplemented with information treatments has a robust consistent effect on affective polarization.

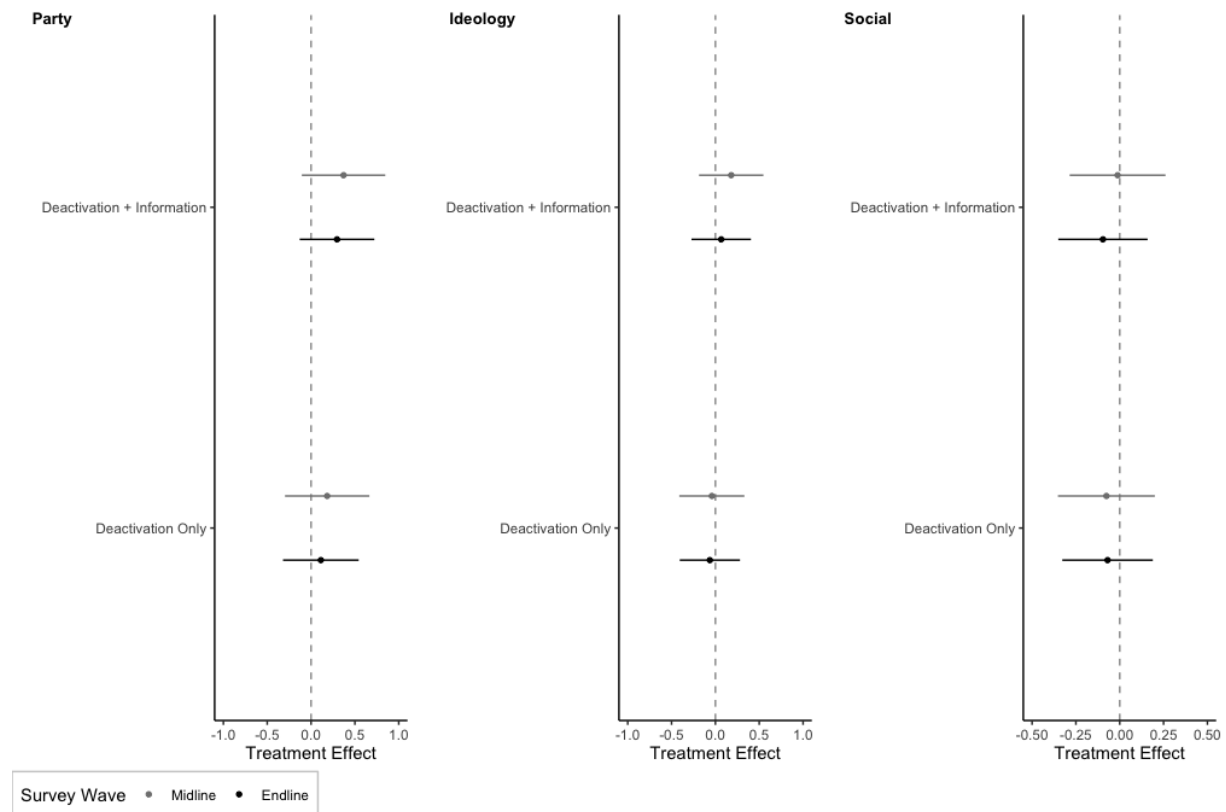


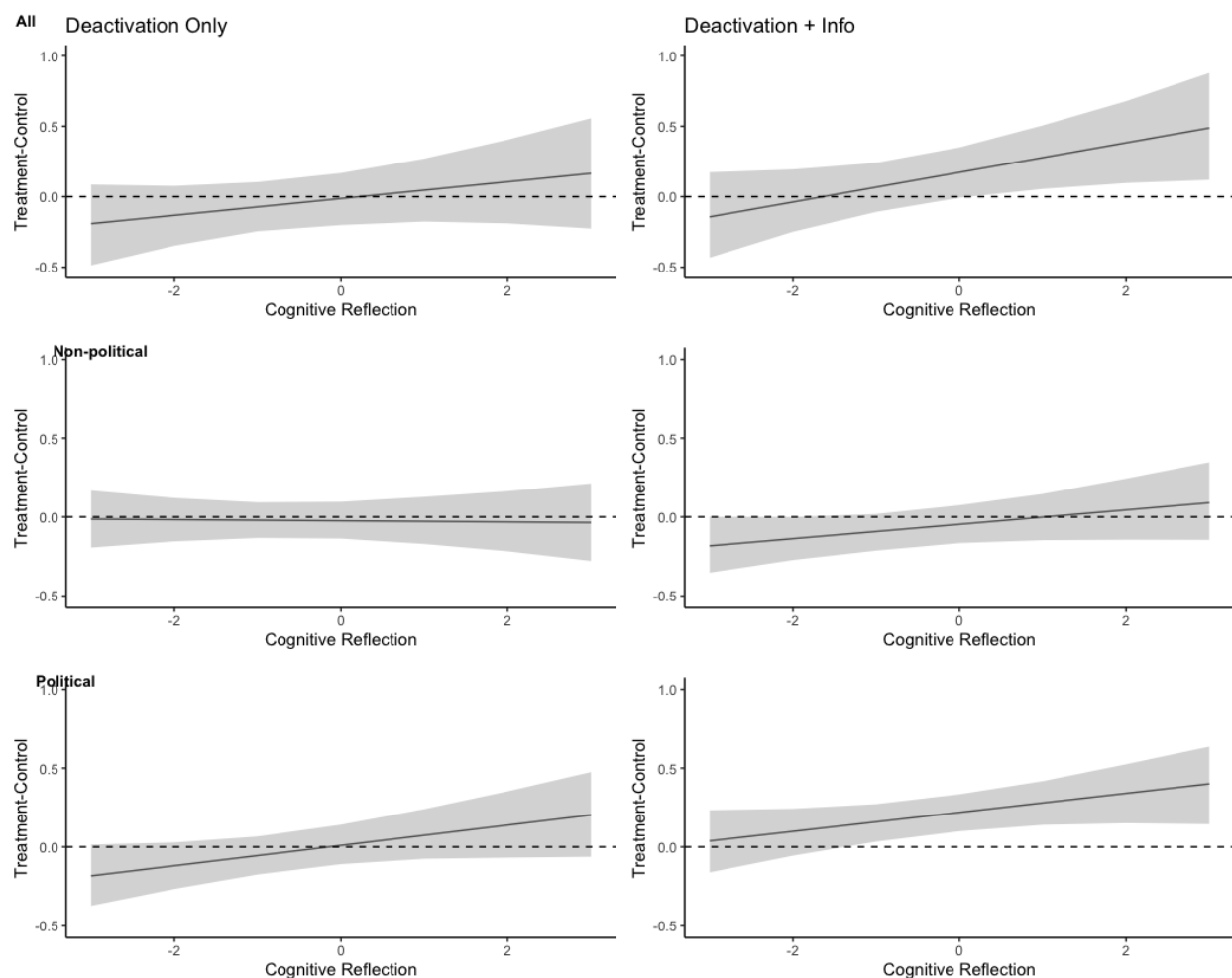
Figure 6: The Effects of Facebook Deactivation on Political and Social Affective Polarization

Note: The dots indicate the difference between the treatment and control group and horizontal lines represent 95% confidence intervals.

Heterogenous Treatment Effects

We preregistered that people’s propensity to second guess their intuitions, known as cognitive reflection (Stanovich 2011), would facilitate the effectiveness of informational treatments and, in doing so, increase political knowledge and decrease affective polarization. These hypothesized heterogeneous effects are consistent with research on the political effects of individual differences in reflection (Arceneaux and Vander Wielen 2017; Pennycook and Rand 2019). We measured cognitive reflection using a three-item Cognitive Reflection Task developed by Frederick (2005) and refined by Thomson and Oppenheimer (2016), and we estimated heterogeneous treatment effects by interacting the treatment indicators with the cognitive reflection task measure. See the SI for question wording and regression results.

Figure 7 offers suggestive evidence that higher levels of cognitive reflection reversed the negative effects of Facebook deactivation on political news knowledge, but only in the Deactivation + Information condition. The x-axis in Figure 7 represents the range of values on the Cognitive Reflect scale: -3 (a participant only provides intuitive but incorrect answers) to +3 (a participant provides all correct answers). The y-axis reported the treatment effect (treatment mean – control mean), with the graphs in the lefthand column reporting the effects of the Deactivation Only treatment and the righthand column reporting the effects of the Deactivation + Information treatment. The first row reports the treatment effects across values of cognitive reflection for the overall news knowledge measure; the second row reports treatment effects across values of cognitive reflection for non-political news knowledge; and the third row does so for political news knowledge.



#E

Figure 7: The Effects of Facebook Deactivation News Knowledge, by Cognitive Reflection (Midline Survey Wave)

Note: The lines indicate the difference between the treatment and control group across levels of cognitive reflection, the dotted horizontal line references zero (no effect), and the shaded area represents the 95% confidence intervals.

We focus on the midline survey, because it is the only wave where we observed a positive overall effect. Consistent with the direct effects reported above, Facebook deactivation had no effect on non-political news knowledge. The Deactivation + Information treatment increased political news knowledge, but these positive effects only reach statistical significance among those with

higher levels of cognitive reflection. We note that, while this pattern of results is consistent with our preregistered hypothesis, it is only suggestive because the interaction between cognitive reflection and the Deactivation + Information treatment is not statistically significant at the pre-registered 0.05 level ($p = 0.157$).

We find a similar suggestive pattern with respect to how cognitive reflection moderates the Deactivation + Information effect on affective political polarization (see Figure 8). In the interest of space, we report the results for the midline survey (when the deactivation protocol would have been in full force), but we find similar results for the endline survey (see SI). The Deactivation + Information condition appeared to *increase* political polarization among intuitive reasoners and this effect attenuates to a null effect among reflective reasoners. These findings are consistent with the thesis that intuitive reasoners are more likely to engage in partisan motivated reasoning (Arceneaux and Vander Wielen 2017). Yet once again they are only suggestive because the interaction between cognitive reflection and the Deactivation + Information treatment is not statistically significant ($p = 0.207$ for party polarization and $p = 0.164$ for ideological polarization).

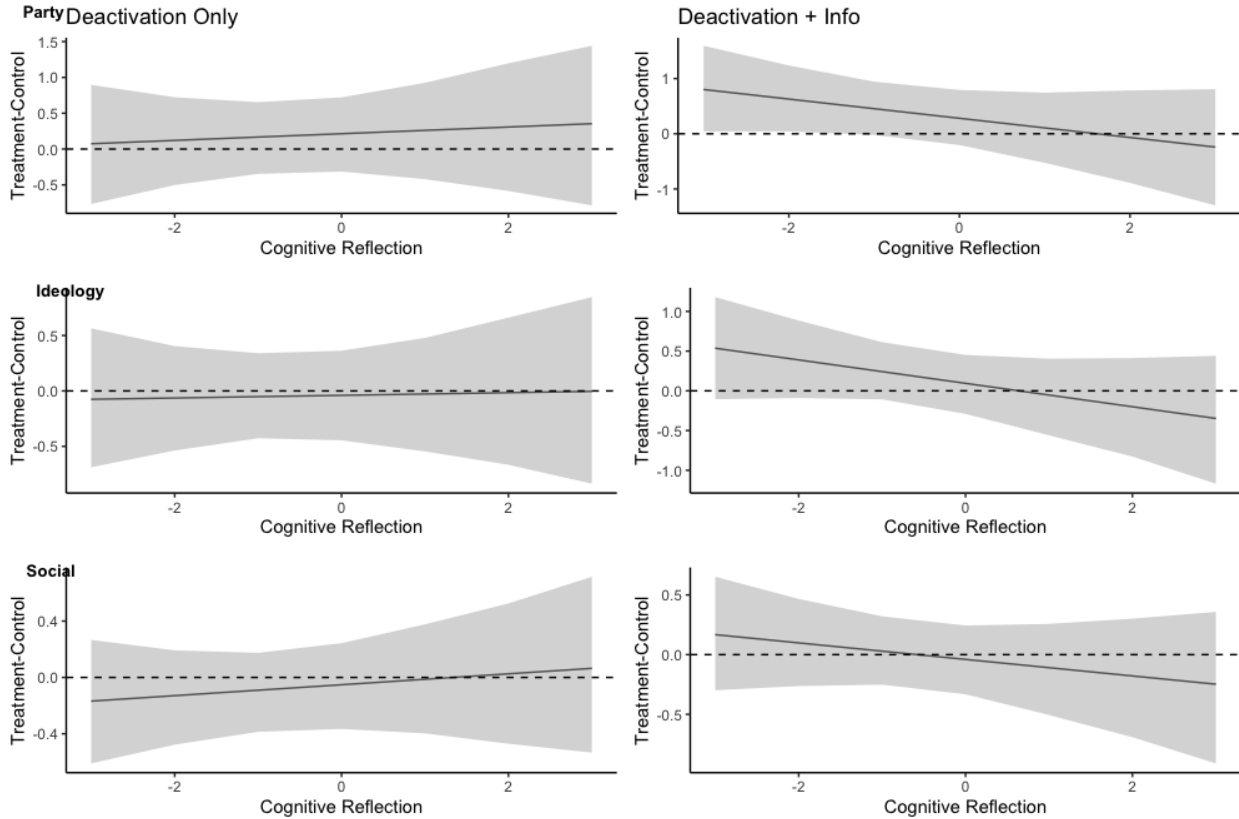


Figure 8: The Effects of Facebook Deactivation on Political and Social Affective Polarization, by Cognitive Reflection (Midline Survey Wave)

Note: The lines indicate the difference between the treatment and control group across levels of cognitive reflection, the dotted horizontal line references zero (no effect), and the shaded area represents the 95% confidence intervals.

In addition to the heterogeneous effects of cognitive reflection, we also preregistered the hypothesis that education would moderate the effects of Facebook deactivation on affective polarization. Following recent research documenting growing political polarization in western countries, including France, between individuals who have a college education and those who do not (Abou-Chadi and Hix 2021; Gethin, Martínez-Toledano, and Piketty 2022), we hypothesized that Facebook deactivation would decrease partisan polarization more among college educated participants than among non-college educated participants. In contrast, we hypothesized that

Facebook deactivation would decrease social polarization the most among non-college educated participants relative to college educated ones. Our reasoning was that social media serves to polarize college educated people along partisan lines, while it does so among non-college educated people with respect to immigration. Our results are summarized in Figure 9. Again, we save space by reporting the results for the midline survey, but we find similar results for the endline survey (see SI). The results corroborate our hypothesis with respect to partisan affective polarization. College educated participants reported lower levels of partisan affective polarization relative to non-college educated participants in both treatments. The data do not support our hypothesis for social polarization, as we observe no statistically or substantively significant difference between college and non-college educated participants' feelings towards those with African ancestry.

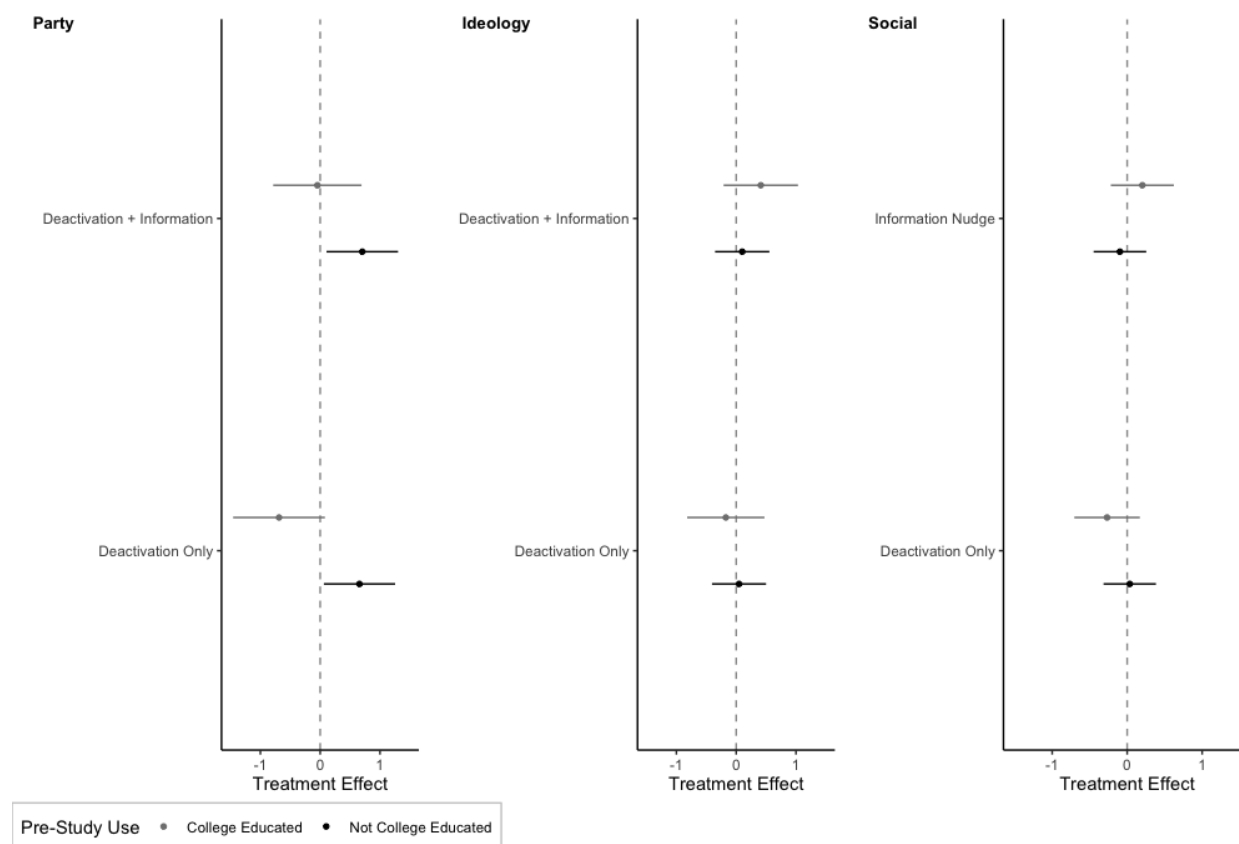


Figure 9: The Effects of Facebook Deactivation on Political and Social Affective Polarization, by Education (Midline Survey)

Note: The dots indicate the difference between the treatment and control group and horizontal lines represent 95% confidence intervals.

Although we did not preregister any hypotheses about the moderating effects of pre-study Facebook use, we replicated Alcott et al.’s analysis where they compared their treatment effects for “heavy” and “light” users. Following their definition, we categorized participants as heavy users if their reported use of Facebook on the baseline survey was above the median and as light users if their pre-study Facebook use was below the median. Like Alcott et al, we find no evidence that the treatments were consistently different for heavy and light users. See the SI for details.

Discussion

Our study replicates and extends previous research on the effects of social media on well-being, news knowledge, and affective polarization. Building on deprivation field experiments conducted in the United States (Alcott et al. 2020) and Bosnia Herzegovina (Asimovic et al. 2021), we incentivized over 1,100 French citizens to deactivate their Facebook accounts during the 2022 presidential election. Our results corroborate previous research with respect to well-being and news knowledge. Participants who gave up Facebook for roughly one month were happier, but less informed about politics than those randomly assigned to the control group who continued to have access to Facebook. In contrast to previous work, which has found that Facebook deactivation can either increase or decrease affective polarization, we found that those who deactivated Facebook were, on average, just as polarized along political and social dimensions.

Another contribution of this study is that it is the first work to combine the parallel stream of research that shows that subtle informational treatments can motivate people to pay more attention to accuracy when evaluating news headlines (e.g., Pennycook, et al. 2021) with the deprivation field experimental design. We find modest evidence that our informational treatments helped people be more accurate when evaluating political news headlines. We only found positive results in the midline survey and these effects dissipated by the endline survey. Moreover, to the extent that our informational treatments were effective, they did so by boosting participants' knowledge about true news, as opposed to enhancing their ability to identify fake news. Beyond news knowledge and contrary to our preregistered expectations, we find little evidence that informational treatments aimed at encouraging people to be civil and recognize the addictive nature of social media caused participants' to be happier than participants who deactivated their Facebook account but did not receive informational treatments, nor did we find evidence that these informational treatments helped reduce affective polarization. Nonetheless, we did find suggestive evidence that our informational treatments may be more effective among those who engage in cognitive reflection (Arceneaux and Vander Wielen 2017; Erlich et al. 2023; Pennycook and Rand 2019). While it is certainly possible that "better crafted" informational treatments would have "worked" for everyone, we note that our informational treatments were no different in terms of quality or content than the those employed in survey experiments. Our findings also fall in line with recent research that call into question the generalizability that subtle informational treatments can consistently motivate behavioral change (Szasz et al. 2022). Overall, our results suggest that claims about the effectiveness of these types of information treatments should be treated with caution. Our preregistered expectations about the effects of these informational treatments almost universally produced null results.

We believe that reproducing previous research findings regarding subjective well-being and news knowledge is important for two reasons. First, it bolsters confidence in the generalizability of the claim that social media use makes people less happy but also more informed about politics. These findings appear to be true across different cultural contexts as well as time. Second, it bolsters our confidence in the integrity of the Deprivation + Information treatment. To the extent we find null effects when we go beyond previous work, we are less concerned that our failure to administer the treatments properly is to blame. Replication also provides clarity on the substantive impact of interventions. In this respect, we find smaller effect sizes than previous research, which reported effect sizes between 0.10 SD to 0.25 SD, whereas we found effect sizes between 0.05 SD and 0.08 SD. Our effect sizes are consistent with the “smallest effect size of interest” in terms of people recognizing that they feel slightly better than before (Anvari and Lakens 2021). In terms of substantive importance, then, our results suggest that if people were to give up Facebook without any incentives, it would potentially make the world a slightly happier and slightly less political informed place.

What should we make of the fact that, across three diverse cultural contexts and different research teams, we observe a consistent finding that giving up Facebook causes people to be less informed about politics? First, it offers strong evidence that Facebook informs people about politics. Despite concerns that social media is awash in fake news and other forms of misinformation about politics, people learn more true things about politics from Facebook than false things. Second, these findings are consistent with the notion that social media, and Facebook in particular, provide political information to people who do not actively seek out the news. Before the rise of social media, the expansion of entertainment choices in the media environment allowed people who were not very interested in news to avoid it, or at least not actively seek it out, which increased the gap in political knowledge between news-seekers and entertainment-seekers (Prior 2007) and blunted the

reach and effects of the news media (Arceneaux and Johnson 2013). The social nature of media, though, means that people are often exposed to information, including information about politics, that they may not seek out (e.g., Settle 2018). If their friends are posting about a political event, such as an election, they will learn about it even if they do not regularly read the newspaper or watch news on television, for instance. In this way, social media allows news to find people (Anspach 2016; de Gil de Zúñiga, Weeks, and Ardèvol-Abreu 2017). Our findings confirm that among the group of people who depend on news to find them, giving up Facebook impoverishes their knowledge about politics. In short, Facebook informs.

At the same time, the informational gains about politics from Facebook appear to come at the cost of reducing people's happiness. Our results, in line with previous research, show that giving up Facebook modestly increases people's subjective well-being. When we dig a bit deeper into the endline survey, 27.6% of participants in the treatment arm reported that they did not miss interactions on Facebook "at all," and only 31.4% indicated that they missed interactions somewhat to very much. Likewise, 36.5% of participants reported that they did not miss sharing content at all, while only 25% reported that they missed sharing content on Facebook. These descriptive statistics suggest that the people's experience on Facebook – the interactions that they engage in or observe, as well as the social engagement that Facebook affords – do not spark joy among most Facebook users. If that is the case, what brings Facebook users coming back to the platform? The answer is, in part, the content: 55.5% of participants reported that they missed the content on Facebook, while only 14% said that they did not miss the content at all. Taken together, these self-reports from the treatment group suggest that the Facebook environment makes people a bit less happy, and yet many choose to come back to the platform in order to take advantage of the unpredictable rewards that it provides.

Contrary to our preregistered expectations, giving up Facebook did not decrease levels of either political or social polarization. These findings are also inconsistent with previous research that found that giving up Facebook decreased political polarization (Alcott et al. 2020) but increased social polarization (Asimovic et al. 2021). While we can only speculate about why we observe these null findings, we find it instructive that giving up Facebook increased political polarization among those lower in cognitive reflection as well as among those who did not possess a college education. These findings suggests that the effects of Facebook on political polarization is contingent on the individual characteristics of the Facebook user. As a result, the general population effect of Facebook on affective polarization may depend on who tends to use the platform as well as the content that they encounter.

Our research helps clarify some aspects of Facebook's societal effects. It informs and disheartens. Future avenues of research could better understand the mechanisms for how and why Facebook informs its users about politics, despite the presence of mis- and disinformation. We also find little evidence that minimalistic informational treatments do much to help counteract the negative aspects of Facebook use, suggesting that we need more theoretically informed work to better understand when and how informational treatments work. A clear limitation of our research is that we focus on one particular social media platform, which has a particular set of affordances and population. While Facebook is one of the most widely used social media platforms in the world, we should be cautious making inferences to other social media platforms with different software and user bases. Finally, our research, like previous research, focuses on an environment in the Northwestern hemisphere, and we should be cautious when it comes to making inferences about the influence of social media in other regions of the world.

References

- Abou-Chadi, Tarik, and Simon Hix. 2021. "Brahmin Left versus Merchant Right? Education, Class, Multiparty Competition, and Redistribution in Western Europe." *The British Journal of Sociology* 72(1): 79–92.
- Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110, no. 3: 629–76.
<https://doi.org/10.1257/aer.20190658>.
- Anspach, Nicolas M. 2016. *The Facebook Effect: Political News in the Age of Social Media*. Dissertation, Temple University.
- Anspach, Nicolas M, Jay T Jennings, and Kevin Arceneaux. 2019. "A Little Bit of Knowledge: Facebook's News Feed and Self-Perceptions of Knowledge." *Research & Politics* 6(1): 205316801881618–19.
- Anvari, Farid, and Daniël Lakens. 2021. "Using Anchor-Based Methods to Determine the Smallest Effect Size of Interest." *Journal of Experimental Social Psychology* 96: 104159.
- Arceneaux, Kevin, and Ryan Vander Wielen. 2017. *Taming Intuition: How Reflection Minimizes Partisan Reasoning and Promotes Democratic Accountability*. Cambridge: Cambridge University Press.
- Asimovic, Nejlá, Jonathan Nagler, Richard Bonneau, and Joshua A. Tucker. "Testing the Effects of Facebook Usage in an Ethnically Polarized Setting." 2021. *Proceedings of the National Academy of Sciences* 118, no. 25: e2022819118. <https://doi.org/10.1073/pnas.2022819118>.
- Bail, C., 2022. *Breaking the social media prism: How to make our platforms less polarizing*. Princeton, NJ: Princeton University Press.
- Bakshy, E., Messing, S. and Adamic, L.A., 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science*, 348(6239), pp.1130-1132.

- Baym, Nancy K., Kelly B. Wagman, and Christopher J. Persaud. 2020. "Mindfully Scrolling: Rethinking Facebook After Time Deactivated." *Social Media + Society* 6, no. 2: 205630512091910. <https://doi.org/10.1177/2056305120919105>.
- Cacciatore, Michael A. et al. 2018. "Is Facebook Making Us Dumber? Exploring Social Media Use as a Predictor of Political Knowledge." *Journalism & Mass Communication Quarterly* 95(2): 404–24.
- Chambers, C., 2017. *The seven deadly sins of psychology*. Princeton, NJ: Princeton University Press.
- Eady, Gregory et al. 2019. "How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data." *SAGE Open* 9(1): 215824401983270.
- Erlich, Aaron, Calvin Garner, Gordon Pennycook, and David G. Rand. 2023. "Does Analytic Thinking Insulate Against Pro-Kremlin Disinformation? Evidence From Ukraine." *Political Psychology* 44(1): 79–94.
- Fletcher, Richard, Craig T. Robertson, and Rasmus Kleis Nielsen. 2021. "How Many People Live in Politically Partisan Online News Echo Chambers in Different Countries?" *Journal of Quantitative Description: Digital Media* 1. <https://journalqjd.org/article/view/2585> (March 14, 2023).
- Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19(4): 25–42.
- Gethin, Amory, Clara Martínez-Toledano, and Thomas Piketty. 2021. "Brahmin Left Versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948–2020." *The Quarterly Journal of Economics* 137(1): 1–48.
- Gil de Zúñiga, Homero, Brian Weeks, and Alberto Ardèvol-Abreu. 2017. "Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics." *Journal of Computer-Mediated Communication* 22(3): 105–23.

- Groenendyk, Eric, and Yanna Krupnikov. 2021. "What Motivates Reasoning? A Theory of Goal-Dependent Political Evaluation." *American Journal of Political Science* 65, no. 1: 180–96. <https://doi.org/10.1111/ajps.12562>.
- Kubin, Emily, and Christian von Sikorski. 2021. "The Role of (Social) Media in Political Polarization: A Systematic Review." *Annals of the International Communication Association* 45(3): 188–206.
- Lee, Changjun, Jieun Shin, and Ahreum Hong. 2018. "Does Social Media Use Really Make People Politically Polarized? Direct and Indirect Effects of Social Media Use on Political Polarization in South Korea." *Telematics and Informatics* 35(1): 245–54.
- Levy, Ro'ee. 2021. "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review* 111(3): 831–70.
- Pennycook, Gordon, and David G Rand. 2019. "Lazy, Not Biased_ Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning than by Motivated Reasoning." *Cognition* 188: 39–50.
- Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles, and David G. Rand. 2021. "Shifting Attention to Accuracy Can Reduce Misinformation Online." *Nature*. <https://doi.org/10.1038/s41586-021-03344-2>.
- Rathje, Steve, Jay J. Van Bavel, and Sander van der Linden. 2021. "Out-Group Animosity Drives Engagement on Social Media." *Proceedings of the National Academy of Sciences* 118(26): e2024292118.
- Ruggeri, Kai, Bojana Većkalov, Lana Bojanić, Thomas L. Andersen, Sarah Ashcroft-Jones, Nélida Ayacaxli, Paula Barea-Arroyo, et al. 2021. "The General Fault in Our Fault Lines." *Nature Human Behaviour* 5, no. 10 (April 22, 2021): 1369–80. <https://doi.org/10.1038/s41562-021-01092-x>.

- Settle, J.E., 2018. *Frenemies: How social media polarizes America*. Cambridge: Cambridge University Press.
- Szaszi, Barnabas, Anthony Higney, Aaron Charlton, Andrew Gelman, Ignazio Ziano, Balazs Aczel, Daniel G. Goldstein, David S. Yeager, and Elizabeth Tipton. 2022. “No reason to expect large and consistent effects of nudge interventions.” *Proceedings of the National Academy of Sciences* 119 (31): e2200732119.
- Thomson, Keela S., and Daniel M. Oppenheimer. 2016 “Investigating an alternate form of the cognitive reflection test.” *Judgment and Decision Making* 11 (1): 99-113.
- Van Bavel, Jay J. et al. 2021. “How Social Media Shapes Polarization.” *Trends in Cognitive Sciences* 25(11): 913–16.
- Zhao, Y., 2014. New Media and Democracy: 3 Competing Visions from Cyber-Optimism and Cyber Pessimism. *Journal of Political Sciences & Public Affairs*, 2(1), pp.114-118.