Growing Closer or Further Apart: Exposure to Social Media in Post-Conflict Settings

Nejla Asimovic[a,e], **Jonathan Nagler**[a,e], **Richard Bonneau**[c,d,e], **and Joshua A. Tucker**[a,b,e]

[a]Department of Politics, New York University, New York, NY 10012; [b]Department of Russian and Slavic Studies, New York University, New York, NY 10012; [c]Department of Biology, New York University, New York, NY 10012; [d]Department of Computer Science, New York University, New York, NY 10012; [e]Center for Social Media and Politics, New York, NY 10012

Despite the belief that social media is altering intergroup dynamics – bringing people closer or further alienating them from one another – the impact of social media on inter-ethnic attitudes has yet to be rigorously evaluated, especially within areas with tenuous inter-ethnic relations. We report results from a randomized controlled trial in Bosnia and Herzegovina (BiH), exploring the effects of exposure to social media during a week of genocide remembrance in July 2019 on a set of inter-ethnic attitudes of Facebook users. We find evidence that, counter pre-registered expectations, people who deactivated their Facebook profiles report *lower* regard for ethnic out-groups than those who remained active. Moreover, we present additional evidence suggesting that this effect is highly conditional on the composition of one's offline environment. We also extend the analysis to include measures of subjective well-being and knowledge of news. Here, we find that Facebook deactivation leads to an increase in subjective well-being and a decrease in the knowledge of current events, replicating results from recent research in the U.S. in a very different context, thus increasing our confidence in the generalizability of these effects.

Social Media | Ethnic Identity | Bosnia and Herzegovina | Conflict | Networks

Does social media usage lead to greater out-group hostility? This question is being asked with increasing urgency in the context of established democracies, and in particular the U.S., with an eye towards establishing whether there is a relationship between social media usage and increasing levels of political polarization (1–3), and, in particular, affective polarization (4). Although most research on this topic has been conducted in the context of advanced democracies, there is an increasing acknowledgment of the need to understand the role social media plays within different political contexts (1, 5), due to the fact that the use of social media and the internet continues to grow in emerging and developing countries (6). Moreover, we know little about the impact of social media usage on attitudes towards ethnic out-groups, as opposed to supporters of opposing political parties.

Despite great interest in the relationship between social media usage and polarization, to date we are aware of only one prior study, carried out in the context of the U.S. and focusing on attitudes towards out-group partisans, that identifies a causal link between social media deactivation and reduction in political polarization, also finding a negative but statistically insignificant effect on party affective polarization (2). While there are a few other randomized impact evaluations of Facebook on users' psychological well-being and civic engagement (7–10), none of these have taken place within a post-conflict country nor have they focused on attitudes towards ethnic out-groups. Particularly within settings with a more recent experience of war, understanding the dynamics of group processes is of vital importance. Given that the way these processes take place in online spaces can help or hinder the goal of transforming relations from antagonistic to constructive, social media should be considered and studied as one of the forces influencing the direction of countries' post-conflict paths.

Importantly, both ethnicity and partisanship are potential markers of one's social identity. Given that partisanship is acquired by choice, however, it is a more informative measure of one's worldview than group membership based on more immutable characteristics such as ethnicity or race (11). As such, some argue that people tend to assign larger blame and responsibility to others for their partisanship than for their inborn group affiliations (12). This difference between partisan and other social identities has recently been put forth to explain why intergroup contact, which is shown to reduce prejudice in a rich body of research (13–15), may be less effective at mitigating out-group hostility in the context of partisanship (16). This, however, remains an open question requiring studies that go beyond partisan affiliation, in particular addressing attitudes towards ethnic out-groups, and especially in an online context.

To address the gaps in the literature pertaining to the causal link between time spent on social media and attitudes

Significance Statement

Amid growing belief that social media exacerbates polarization, little is known about the causal effects of social media on ethnic out-group attitudes, particularly within post-conflict societies. Through an experiment in Bosnia and Herzegovina where users refrained from Facebook usage during a week of heightened identity salience, we find that – counter expectations – people who deactivated their accounts reported *lower* out-group regard than the group that remained on Facebook, but that this effect was likely conditional on one's offline environment. Additionally, we replicate findings from a study on U.S. users: deactivation led to a decrease in news knowledge and an increase in subjective well-being. Our findings provide a corrective to assessments, frequently dichotomous and overly simplistic, of the impact of social media on societal dynamics.

N.A. and J.A.T. designed the research; N.A performed the research; N.A., J.N, R.B, and J.A.T. planned the analyses; N.A. analyzed data and wrote the first draft of the paper, and all authors contributed to revisions.

J.A.T. received a small fee from Facebook to compensate him for administrative time spent in organizing a 1-day conference for approximately 30 academic researchers and a dozen Facebook product managers and data scientists that was held at NYU in the summer of 2017 to discuss research related to civic engagement; the fee was paid before any data collection for the current project began. He did not provide any consulting services nor any advice to Facebook as part of this arrangement. N.A., R.B., and J.N. declare no conflict of interest.

[2]To whom correspondence should be addressed. E-mail: nejla.asimovic@nyu.edu.

towards out-groups based on more immutable identity categories such as race and ethnicity, as well as the lack of research on social media's impact in post-conflict societies, we conduct a pre-registered experimental field study of the effects of exposure to social media within the post-conflict setting of Bosnia and Herzegovina (BiH) on *ethnic* out-group attitudes. More specifically, we randomly subset users to deactivate Facebook accounts during a week around the Srebrenica genocide commemoration – a period of heightened attention to past ethnic conflict – and assess inter-ethnic regard as the primary outcome of interest. We find, contrary to our pre-registered hypotheses based on the existing literature, that *decreased* use of Facebook led to *more negative* attitudes about ethnic out-groups. Moreover, a series of supplementary analyses – not pre-registered, but undertaken in an attempt to better understand our surprising finding – suggest that these effects are concentrated predominantly among those who live in ethnically homogeneous environments, i.e. people whose *offline networks* are likely to be ethnically homogeneous (Fig. 3, Table S10-S12). We also extend our analysis to include assessment of the impact of social media usage on knowledge of the news and subjective well-being, in an effort to examine the robustness of prior findings from the U.S. (2). In line with our pre-registered hypotheses based on the prior findings from the U.S.(2), we find that decreased Facebook usage improves subjective well-being but decreases knowledge of current events, thus advancing our confidence in the generalizability of our findings as well as the validity of our experimental set up.

Our research is motivated by the debates around the challenges and opportunities for intergroup contact brought forth by the proliferation of social media. As social media usage has spread across the globe, there has been a great deal of optimism about the potential for social media usage to encourage the sharing of information among different groups (17, 18), as well as to serve as a tool for promoting inter-community relations; and to shift perceptions and behaviors by increasing mutual understanding between antagonistic groups through a platform on which people can engage in discussions even across social identity lines (19–21). Social identities can be activated on social media in a myriad of ways: the overall context of the social media environment; nonverbal and visual cues by users; and even linguistic cues that denote group identity are only few of the possible options (22, 23). By its nature, social media allows for direct access to individual voices which can be voices of either ethnic hatred or ethnic solidarity. As the communication moves from a "one-to-many" to a "many-to-many" structure (1), individual users become creators of content with the opportunity to share their own messages with the wider public, a feature that used to be reserved only for elites and traditional media. Allowing for the individual out-group voices to be heard can also have a personalizing effect, especially within contexts such as the one we study in which much of the official discourse categorizes and draws boundaries between groups of people based on their ethnic membership. Even when users do not reveal their political thoughts online, they will be exposed to the posts their friends (and often friends of the friends) share, comments their friends write, and discussions in which they engage, all of which may be of a different nature than the official rhetoric to which they would otherwise be exposed. Compared to the offline world, social media platforms facilitate more connections with "weak ties", defined as acquaintances that link more distant clusters of people while introducing novel information and more diverse views into the conversation (24). Having weak ties across social cleavages forms the basis for growing what in the literature is referred to as "bridging social capital", which is hypothesized to facilitate cross-ethnic cooperative relationships (25, 26). Such relationships are also enabled by the unique characteristics of online communication — the ability to transcend physical distance, lower the cost of contact, more easily create perception of equal status, and reduce anxiety between members – providing a promising basis for effective intergroup contact (27).

Sharing the same online space with out-group members, however, does not necessarily satisfy the facilitating factors under which contact experiences are more likely to reduce prejudice: equal status, intergroup cooperation, common goals, and support by social and institutional authorities (28). Although electronic interaction more frequently provides the perception of equal status due to status differentials being less observable than within face-to-face contact (27), other conditions are less frequently satisfied as part of one's online experience.* While previous research shows a positive association between Facebook usage and bridging social capital (29, 30), it also reveals a positive association with bonding social capital, which refers to resource sharing with strong ties and homogeneous social networks (i.e., fellow in-group members). Strong levels of bonding social capital, while important for social support, may provide increased opportunities for ethnic entrepreneurs (31), thus endangering democracy in ethnically diverse societies (32). Scholars are also increasingly raising concerns that bridging social capital is endangered in online social networks by platforms' algorithms (33). There is growing evidence suggesting that the way in which platforms' algorithms deliver content may be inducing echo chambers, online incivility, and hate speech, all of which in turn contributes to polarization and further exacerbates societal problems (34–36), which can be particularly problematic within politically fragile post-conflict societies.

We considered these competing arguments and decided to base our pre-registered hypotheses on evidence that when faced with positive and negative content, adults tend to display a negativity bias in the formation of their impressions and general evaluations. The *negativity bias* phenomenon is defined as a human tendency to attribute negative outcomes to the intentions of another person more frequently than similar neutral or positive outcomes (37–39). Within the context of our study, names carry cues as to a person's ethnic membership, which could be used to translate acts from an individual to update beliefs about the whole community. We also followed what is argued in the case of political identities – that the awareness of identities, particularly when heightened, as we posited would be the case during the commemorative period, would increase group polarization (40). Hence, the first hypothesis we pre-registered was:

H1: Users who are not active on Facebook during the week of genocide commemoration will display more amicable inter-ethnic attitudes compared to their counterparts who remain active.

*A meta-analysis of 515 studies of intergroup contact concludes, however, that these conditions increase the chance but are not necessary for contact to produce a positive effect (13).

In addition to testing our expectations around the effect of social media on out-group regard, our research design also allows us to test the robustness of two key findings – that Facebook deactivation improved subjective well-being and decreased knowledge of current events (hereafter "news knowledge") – from the only prior Facebook deactivation study, which as noted previously was not conducted in a post-conflict society but rather in the U.S. (2). While there is research suggesting that social media can have a positive effect on users' levels of psychological distress by enabling them to stay in contact with their extended families, receive more social support, or access health information (41), a growing literature reports the negative psychological impact of the time spent on social media platforms. Such negative impact is attributed to the features of social media platforms that can encourage social comparisons, induce addictive behavior, or enable cyberbullying (42, 43). In pre-registering our own hypothesis, we drew most heavily from the largest-scale causal evidence to date which found that a month-long Facebook deactivation led to small but significant improvements in well-being among the U.S users (2). We did not expect that the post-conflict nature of the context would lead to a different direction for the effect of Facebook deactivation, although it seemed plausible that the effect size, especially as it relates to anxiety levels, would be stronger given the emotionally-charged nature of the commemorative period in which we conducted the study:

H2: Users who are not active on Facebook during the week of genocide commemoration will display higher levels of subjective well-being compared to their counterparts who remain active.

In addition to subjective-well being, we test the effect of Facebook deactivation on news knowledge. In the previously referenced Facebook deprivation study which took place in the U.S. (2), the authors found that deactivation reduced their news knowledge index by 0.19 SD, and speculated that the magnitude of the effect might differ depending on the duration of the time spent deactivated from Facebook. We saw no reason why we would expect this dynamic to play out differently within a post-conflict society. In an attempt to evaluate the external validity of that finding, we therefore test the following hypothesis:

H3: Users who are not active on Facebook during the week of genocide commemoration will display lower level of news knowledge compared to their counterparts who remain active.

We also pre-registered a measure of political disaffection, which we do not focus on in the main text, but for which we provide a detailed analysis within the Supplementary Information (SI, Section 13).

1. Research Design

Our study took place in Bosnia and Herzegovina (BiH) from July 7 until Jul 14 of 2019, encompassing the week around the Remembrance Day (July 11) on which the country commemorates the Srebrenica genocide from 1995, with various events each year including a mass funeral for the victims identified over the previous twelve months. Online and offline discussions related to the war are held to a certain extent throughout the year, but intensify particularly strongly in this period. The genocide in Srebrenica is central and deeply intertwined to the memory of the Bosnian war, as the worst atrocity in Europe after the World War II. Legal battles over war crimes, in the twenty-five years since the end of the war, have been contentious and genocide denialism continues to grow. Despite the fact that International Criminal Court of Justice (ICJ), among other institutions, ruled that the acts against the Bosniak Muslims committed in Srebrenica were acts of genocide under the international law (44), significant number of officials and ethnic entrepreneurs continue to reject such rulings. The commemoration period serves as a natural prime of ethnic identities and, more generally, a proxy for periods during which identities (in this case, ethnic) are made even more salient and more fervently discussed.

Participants were recruited through Facebook advertisements which we ran across BiH (Fig. S1.B) in both Cyrillic and Latin alphabets. Importantly, there is no language comprehension barrier between the members of three main ethnic groups in BiH. Out of the individuals who successfully completed the baseline survey, we selected only those who reported following at least one BiH news or political page on Facebook, and reported being never or almost never on Twitter nor Reddit. We chose these criteria to increase the likelihood of our sample being exposed to political or intergroup content, as well as to reduce the extent to which the discussions on other social media platforms (Twitter or Reddit) might be influencing the outcomes of our interest. Regarding the former, it should be noted that Facebook is by far the largest social media platform in BiH encompassing 99.02% of social media market share, compared to Twitter and Reddit with 0.37% and 0.01% of the market share respectively (45). Hence, active Twitter or Reddit users tend to be a more distinctive group and by excluding them, we focus our attention to an average Facebook user responding to our ad.[†] We used blocked randomization to divide the 556 resulting individuals into treatment and control, blocking on the variable from the baseline survey that measures the importance of ethnic identity to one's self-identification. 258 in the control group and 263 in the treatment group were successfully e-mailed. After we informed individuals of their assigned treatment, 92 people in the treatment group and 64 in the control group either never responded or responded too late to be included as participants. Of those who did confirm their participation after receiving this information, 31 participants (15 from the treatment group and 16 from the control group) failed to complete the study. Since attrition may introduce bias if it is systematically related to the outcomes of interest, we analyze the baseline characteristics of users who did not finish the final survey. We do so using the sample of all users included in randomization; we also present an attrition analysis with the subsample of users who attritted after initially providing the affirmative response (SI, Section C). In neither case do we find significant imbalances between the characteristics of those who attrited after being assigned to deactivate versus those who attrited within the group assigned to remain active.

We received endline surveys from 159 participants in the

[†] For comparison, the data for the US over the same period suggests that the social media market share of Facebook, Twitter and Reddit is 50.93%, 18.17% and 0.55%, respectively.

treatment and 194 participants in the control group, with their observable characteristics balanced between the treatment and the control group (SI, Table S3). The only imbalance detected is on the self-reported weekly frequency of accessing Facebook, with participants in the control group reporting somewhat higher values in the baseline survey compared to the treated group. We control for this covariate in the main results (Model 2), which does not substantially change the estimated treatment effects. The final sample, with descriptive statistics reported in Tables S1-S3, consists primarily of participants who identified themselves as Bosniaks (58.92%), followed by those identifying as Serbs (15.71%) and Croats (6.52%). Finally, 13.03% of respondents chose to identify as Bosnians and 4.82% chose to not report their ethnic identification.[‡]

The deactivation was monitored via a Python script that automatically checked Facebook URLs twice a day and sent a report to researchers with the IDs of participants who remained active. In addition, we manually checked the URLs at random times throughout the day. All the initial URLs were also transformed into numeric IDs, so that we would be able to detect activity from the profile even if a user changed the way her username is shown to the general public. Once deactivated, the group complied with deactivation at a rate that exceeded 98 percent. Those who did not abide by the treatment and activated their profile at some point during those seven days were contacted to describe their reason for activation and provided with a final warning (e-mail and text); if they did not deactivate within 1 hour after the e-mail was sent to them, they were excluded from the study. The control group was asked to continue using Facebook as they regularly would, while also sharing the information on the exact daily amount of time they spent online by sending screenshots of the report that Facebook creates for each user.[§]

Participants took a comprehensive survey at the end of the experiment, and we focus on questions within three families of outcomes: out-group regard, subjective well-being and news knowledge (full list of questions in SI, Table S4). All of the questions were pre-registered; following our pre-analysis plan, we present results both for the composite indices and each of their corresponding indicators.[¶] With the following regression, we estimate the intent-to-treat effect of Facebook deactivation on our outcomes of interest:

$$Y_i = \alpha + \beta T_i + \theta X_i + \epsilon_i, \quad [1]$$

where Y_i is an outcome, $T = \{0, 1\}$ an indicator of treatment assignment, and X_i a vector of covariates for the individual. Full covariate specification includes the following controls captured in the pre-treatment survey: gender, age, employment status, ethnicity, weekly frequency of Facebook usage and the perceived importance of country and ethnic membership to one's identity. When incorporating covariates and in an effort to improve consistency of the estimated effect, we include mean-centered covariates and interact them with the treatment indicator (46, 47).

[‡] According to Census 2013, the composition of ethnic groups in Bosnia and Herzegovina is as follows: 50.11% Bosniaks, 30.78% Serbs, and 15.43% Croats.

[§] In November of 2018, Facebook rolled out a new feature called "Your Time on Facebook" which tracks the amount of time a user spends on the Facebook mobile app. Instructions on how to access this feature were sent to all participants in the control group, yet depending on the device used, not all users were able to access or send the updated reports, and thus self-reported the time instead. We do not use this information in the analysis..

[¶] Deviations from the pre-analysis plan (in terms of final survey measures) are incorporated within Table S4.

2. Results

We first present our results on the impact of Facebook deactivation on subjective well-being and news knowledge, after which we present the effect that deactivation had on users' inter-ethnic attitudes.

A. Main Effects on Subjective Well-Being & News Knowledge.

We test our "news knowledge" variable by creating an 8-item knowledge quiz with news headlines, some of which truly appeared over the seven days of the treatment and some of which were written by our research team to provide a combination of true and false news. Participants were asked to indicate, without checking the information on the internet, whether the headline was true, false or whether they were unsure. The *news knowledge index* was created as a count of correctly assessed statements minus the number of statements for which the respondents gave the wrong response. What we find is that Facebook deactivation significantly reduced news knowledge, with treatment leading to a reduction of news knowledge by 0.27 SD (SE=0.106; $p<0.05$) (Fig. 1: Panel A). This effect is 0.08 SD larger than the effect detected in the study on the U.S. sample. As the authors of the U.S. study discuss (2), a longer period without Facebook is likely to have a lesser impact on news knowledge as users find alternative sources of information, which could be one of the reasons explaining why the magnitude of the effect they detect after a month of deactivation is somewhat lower than the magnitude we find with a shorter treatment. They furthermore propose that one way in which Facebook deactivation might be reducing news knowledge is by making the treatment group participants more likely to answer "unsure" (2). We find a similar pattern within our sample, with the treatment group answering an average of 3.62 statements out 8 questions with "unsure", as compared to an average of 3.29 within the group that remained active on Facebook.

With regards to subjective well-being (Fig 1: Panel A, Table S5), we detect statistically significant effects on levels of anxiety (ITT: β= -0.37 SD, SE =0.108, $p<0.01$) and loneliness (ITT: β= -0.20 SD; SE=0.107; $p<0.10$). Anxiety was also identified as one of the variables with the largest and most significant effects (0.10 SD) in the U.S. deprivation study (2). The reason why the decrease in anxiety level that we detect is much larger could be related to the period in which we conduct the analysis – the week around Srebrenica genocide commemoration – during which the content to which users are exposed online is particularly emotional and distressing. The effect estimate on the index of subjective well-being, created as a sum of z-scores, yields a marginally significant effect size of 0.18 SD (SE=0.105, as shown in Table S5). It should be noted that a one-sided test of the unadjusted difference in means of the composite well-being index, consistent with the direction of our hypothesis and excluding the box-plot outliers, yields a p-value of 0.03[‖]. In the Supplementary Information, we present regression results incorporating different sets of controls (Table S5) and the effects remain stable across specifications. When we adjust the p-value for false discovery rate (48) (Table S7), the coefficient on anxiety remains statistically significant, but not the coefficients for loneliness and the aggregated index of well-being. Taken together, we conclude that deactivation of Facebook led to a strong and significant decrease in users'

[‖] Outliers here are defined as observations standing above and below the 1.5*IQ.

anxiety levels, with positive improvements in other components of users' subjective well-being.

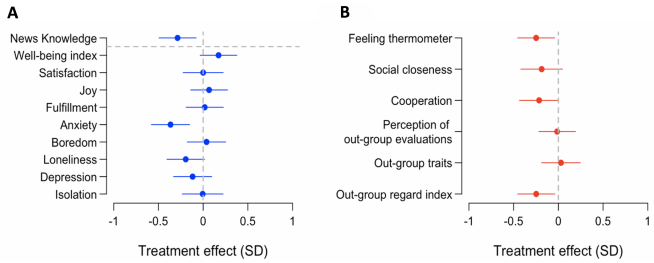

Fig. 1. Intent-to-treat treatment effects of Facebook deactivation estimated on subjective well-being (Panel A), news knowledge (Panel A) and inter-ethnic attitudes (Panel B), with a full set of controls: gender, age, employment status, ethnicity, weekly frequency of Facebook usage and the perceived importance of country and ethnic membership to one's identity. Each coefficient is reported with its corresponding 95% confidence intervals based on robust standard errors, and is standardized relative to the SD of the control group (N=353). In the creation of the well-being index (additive index of z-scores), depression, loneliness, anxiety, boredom and isolation, were reverse coded so that higher values indicate more positive attribution.

B. Main Effects on Inter-Ethnic Regard. Analyzing the impact of the treatment on inter-ethnic attitudes, we find that deactivation from Facebook led to more *negative* rather than positive out-group attitudes, the opposite of what we initially hypothesized. The two outcomes for which we see very little movement include perception of out-group traits (asking subjects to evaluate traits of the out-group members) and perception of out-group evaluations (asking subjects to evaluate how they think their in-group members are perceived by the members of other ethnic groups). However, we find that deactivation reduced the reported willingness to cooperate with the out-group by 0.21 SD (SE=0.112; $p<0.1$), while significantly reducing feeling thermometer score by 0.24 SD (SE=0.105; $p<0.05$), controlling for randomization block and a rich set of baseline covariates. We consider measured indicators to be different ways of capturing a latent variable of out-group regard and, informed by factor loadings, create our main index of out-group regard as a principal component score of the five indicators (feeling thermometer, social distance, willingness to cooperate, out-group trait ratings and perception of out-group evaluations, as described in SI, Table S4). With a principal component index of out-group regard, we find that deactivation significantly reduced reported levels of out-group regard by 0.24 SD (SE=0.105, $p<0.05$). The estimate on the principal component index of out-group regard passes the Benjamini-Hochberg multiple-comparisons correction at the 0.10 level (SI, Table S8). We also tested – and show in Section 5 of the SI – that our results are largely robust to outliers. To understand the magnitude of the effects, it is useful to contextualize the results by putting them within a comparative context. Although there are still no published and comprehensive evaluations of long-term trends in affective polarization across the developing world, it is informative to note that a recent U.S. study (49) shows an increase of 0.72 SD in affective polarization by 2016, measured with feeling thermometer results and taking 1978 as the baseline. An effect size of 0.24 SD is a third of that increase and is, as such, of considerable importance.

The finding that Facebook deactivation had a negative impact on users' out-group attitudes goes against the theoretical insights on which we based our hypothesis, suggesting the need for theory refinement and further exploration of our findings.

3. Discussion

In this section, we move beyond our original preregistered study in an effort to shed some light on our unexpected findings: why might it be the case that moving off of Facebook at the time of a genocide commemoration led to *higher* levels of out-group animosity? While a full consideration of this topic would require new preregistered studies to address this research question specifically, for now we use data from our original experiment – as well as some additional data we collected expressly for this purpose – to provide an exploration of the impact of one potential explanation: differences between the composition of one's offline and online networks, especially as it relates to the degree of exposure to the out-group.

As our starting point, we acknowledge that, in interpreting the role of any media on shaping social dynamics, focusing solely on direct media exposure is insufficient without taking into account the alternative activities that such exposure is crowding out (i.e. direct versus substitution effects) (50–52). Intriguingly, when asked how they spent their newly freed-up time, the treated group (deactivated from Facebook) reported spending more time than usual with friends and family as the most frequent response (Fig. 2); further analysis revealed that the treatment of Facebook deactivation increased the amount of time users spent with friends and family, as well as the amount of time spent reading the news on other online sources, at $p<0.06$ significance level (Fig. S3).

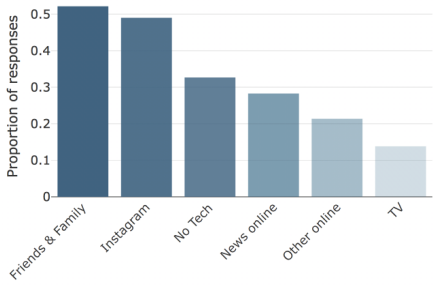

Fig. 2. Histogram of the treated group's responses (N=159) to the final survey question (multiple responses permitted): "In the last week, relative to what is typical for you, would you say you spent more or less of your free time..."

Given the commemorative period during which this study was conducted, it is important to note that people exhibit a tendency to share emotional experiences (53) and discuss contentious events, such as wartime-related topics and politics, with those with whom they feel closest. Scholars of social psychology explain how experiences of collective trauma and suffering feed back into the present, with narratives of family members often cited as the most powerful influences in transmitting and forming awareness of group victimization (54). Moreover, some scholars (55, 56) have suggested that vulnerability to echo chambers may be greatest in offline social networks, and point out that we have lost sight of the fact that offline social networks are oftentimes even more homogeneous than the online networks.

The lack of exposure to the out-group and segregation – which tend to be the enduring features of post-conflict settings – are associated with a plethora of discriminatory behaviors and attitudes. To alleviate those negative effects, scholars propose and show that contact with individual out-group members can translate into lower prejudice levels toward the entire group (14), with positive association between contact and prejudice reduction supported in the seminal meta-analysis of 515 intergroup contact studies (13). Previous research also illustrated that even those who experience no direct contact with the out-group may benefit from living within a diverse setting with in-group members who do partake in such contact (57). This contextual effect can translate to online platforms as well, given that social media provides easier access to the extended networks of users' online connections, which may introduce novel information, especially for the dyads of Facebook friends that belong to non-overlapping social circles. Our treatment of social media deactivation which reduced, if not entirely eliminated, participants' contact with the out-group, may have made some users more dependent on offline echo chambers and moved them further away from the opportunity to engage with individual voices, as opposed to solely having the effect of reducing exposure to online divisive rhetoric. For those who have no contact with the out-group as part of their offline experiences either, the contact hypothesis could therefore predict that the treatment of deactivation would have a negative effect on their out-group regard.

To explore whether this implication holds in the context of our study, we performed a series of analyses that were not explicitly part of our original pre-registration, and thus should be interpreted with that caveat in mind.** To proxy for users' offline networks, we use the latest Bosnia and Herzegovina Census data (2013) on ethnic heterogeneities of BiH localities. This proxy is suitable as research has shown that county-level political and racial heterogeneity can help predict individual-level heterogeneity of political discussions, which we posit could hold for ethnic heterogeneity as well (58). Our conceptualization of offline networks goes beyond friends and family of users to include the composition of individuals with whom users might come into contact, both in the workplace and in daily activities. Measuring diversity is complex, with a variety of indices employed depending on the theoretical questions of interest. We test and check the robustness of our insights using three of indices with which ethnic heterogeneity can be quantified (formulas and detailed descriptions of the indices in SI, Section 9): an index of ethno-linguistic fractionalization, calculated using the Herfindahl concentration index, which is a widely employed measure of ethnic diversity; the share of the majority group, a measure that takes into account the size of the largest ethnic group in a particular town following the information from the latest Census; and the Shannon diversity index, a mathematical measure most frequently used to characterize species diversity within a community. For each of these three diversity indices, we follow the same procedure. We first calculate the heterogeneity index of users' networks and then subset the data into below and equal or above the median of the index value, thus categorizing towns into more and less heterogeneous. We then rerun our original analysis within the two sub-samples, and evaluate whether the effect of deactivation on out-group regard is more negative for users who live in more homogeneous communities. Indeed, this is exactly what we find using all three indices of ethnic diversity (Fig. 3): within the subsample of more homogeneous towns, the effect of deactivation on the composite index of out-group regard is negative and highly statistically significant (SI, Tables S10-S12); on the contrary, for people living in more heterogeneous offline environments, there is no effect of deactivation on out-group attitudes.

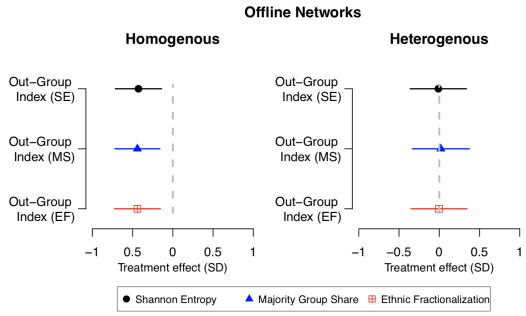

Fig. 3. Subgroup analysis: Treatment effect (full covariate adjustment) of Facebook deactivation on the composite index of out-group attitudes for users with offline networks equal or above and below the median value of heterogeneity, as measured by the three indices of ethnic diversity (N=172, N=177, N=176, for Shannon entropy, majority group share and ethnic fractionalization measures respectively).

Hence, it appears that the negative effect of Facebook deactivation on out-group regard is almost entirely driven by users within the communities in which the opportunities for offline intergroup contact are limited (and in some cases possibly non-existent), and therefore may be living within offline echo chambers that are stronger than the ones users find online. Although smaller sample size can lead to overestimating the magnitude of subgroup effects (59), we find consistent conclusions across different ways of measuring the effect. While the subgroup results in Figure 3 rely on a binary indicator, the moderation analysis is robust to using a continuous measure of town heterogeneity (SI, Tables S16-17). Across all three methods of measuring diversity, the estimate on out-group regard passes the Benjamini-Hochberg multiple-comparisons correction at the 0.05 level (SI, Tables S13-15).

For the heterogeneity of online networks, we invited interested participants to download and share with us the list of friends with whom they are connected on Facebook. We gathered this information from 134 participants, which constitutes 38% of our initial sample.†† We simultaneously created a comprehensive database of names categorized by likely ethnicity, which enabled us to obtain percentages of Bosniak, Serb, and Croat names in the networks of each of our users, as identified from the three categories of names in our database. On average, we were able to estimate the ethnicity of 74.9% of the names within the online network of each user. Manual inspection of the other names suggests that most unmatched names mainly belong to one of the three categories:

**We did, however, pre-register that we would map the level of ethnic heterogeneity within users' online networks from their list of Facebook friends (presented later in the section); and that we would examine how the level of segregation affects ethnic divisions, so these new analyses are very much in the spirit of that original plan.

††It was a pleasant surprise that this many people elected to share their list of friends with us, especially given the effort involving in downloading this data from Facebook, but also of course far short of full compliance. Therefore, results from analyzing these data should be considered especially exploratory. We compare baseline characteristics of those who shared versus those who chose not to share their online data in the Appendix, Section 11, and find no significant imbalances. It is plausible, of course, that the groups differ on an unobservable characteristics, e.g. trust toward researchers, that our baseline survey measure does not capture.

non-Bosnian names (possibly friends from other countries), nicknames (unique combinations of letters), and non-human accounts (coffee shops, artists, and locations.) Following the same steps as for the analysis of offline networks, the direction of the estimates suggests that the negative effect of deactivation on out-group regard (SI, Table S19-S20) is decreasing in homogeneity of the online network (which, as expected, contrasts with the increasing, i.e. becoming more negative, effect in homogeneity of offline network described previously). Although the estimated interaction effect between the heterogeneity of one's online network and the deactivation treatment is in the expected direction, the standard errors are too large to draw firm conclusions. It should be noted that this measure captures only users' direct outgroup friends, and is likely an underestimation of the total exposure to the outgroup as it overlooks exposure through one's extended networks (friends of friends). And indeed, there is some evidence suggesting that such exposure may be substantial within the Balkan region. In 2016, Facebook introduced the first comprehensive measure of social connectedness across the world, constructing a dataset of aggregated information of friendship links between all Facebook users (60). Researchers have found that the level of connectedness is higher between regions who formed part of the same country in the past, and highlight some regions in the former Yugoslavia as being several times more connected compared to similar pairs of other European regions (60).

We also use the available data to test whether those users whose online networks are more diverse than their offline networks are more negatively affected by the treatment of deactivation. Again, this analysis should be considered exploratory in light of the fact that it was not pre-registered, yet we believe it suggests an important difference in the way Facebook deactivation affects people whose online and offline network diversity metrics differ. Using the same measures – ethnic fractionalization, Shannon entropy, and majority group share – we estimate the diversity of users' online networks, and create a binary variable that indicates users whose online networks are more diverse than their offline ones.[‡‡] The interaction term between the treatment and this indicator reaches statistical significance with out-group regard formed as a sum score and diversity measured with any of the three indices (SI, Fig. S7 and Table S21), while confidence intervals are wider when the index of out-group regard is created as a principal component score. Across the two operationalizations, however, the direction of the estimates suggests that deactivating from social media had a comparatively more negative effect on out-group attitudes for users whose online networks are more diverse than their offline ones, which we are suggesting occurred because the deactivation likely decreased exposure of such users to the out-group. It is important to note that our results are aligned with two possibilities: that those who stayed active on Facebook experienced an improvement of out-group attitudes because of their online contact and discussions engendered by the remembrance period; or, alternatively, that those who deactivated their Facebook accounts experienced worsening of their out-group attitudes because they were primarily exposed to discussions within homogeneous offline networks or the official discourse.

Theoretically, there is a potential, alternative, non-network based, explanation for our findings: that Facebook content about the commemoration promoted reconciliation, in which case the results we observe could be a product of the positive skew in the news content on the platform, irrespective of the homogeneity of either one's offline or online networks. To gauge the type of political content that was circulated during this period, we conduct a qualitative content analysis of the top five Facebook news pages in terms of the number of followers and – when these two categories do not overlap – Facebook pages of the top five news outlets in terms of the average number of website visitors per month. We find little to no evidence of unifying campaigns or promoted content, and, in fact, find the contrary to be the case in our sample: most articles are emphasizing group divisions (for more, see SI Section 1). The same, if not more, is true in the case of print media and public broadcasters, where, despite the existence of multiple media sources, pluralistic views remain rare (61). Of course, it is possible that those in the deactivated treatment group could have accessed different media articles about the commemoration through means other than Facebook. This, however, does not explain the observed heterogeneity in the effects depending on the composition of users' networks. Moreover, media outlets increasingly rely on mobile resources to share the news and distribute information through social networks, with Facebook being the most frequent option (61). As such, many news outlets tend to promote their content through both their official websites and through Facebook. Since a user could presumably access similar, if not the same, news in print, through a website or distributed via social media, we argue that it is the experience of engaging with such content that makes social media uniquely different, given the opportunity to participate in discussions, as well as to observe one's social network and larger community engage with the online content.

4. Conclusion

Restoring social capital after conflict in societies with tenuous inter-ethnic relations, as well as maintaining constructive inter-ethnic relationships within a multi-ethnic society, is a complex and ever-evolving process. The rise of digital technologies and the increase in social media penetration across the world has for many changed the environment within which group processes unfold, yet we still lack rigorous evidence about the consequences of these developments. Going beyond the U.S. context and partisan attitudes, we provide causal evidence of deactivation from Facebook leading to an unexpected *decrease* in users' regard toward ethnic out-groups. In addition, we find preliminary evidence suggesting that such impact is conditional on the composition of users' offline environments. We also find that recent findings from the U.S. (2) of a positive effect of social media deprivation on subjective-well being and a negative effect on levels of news knowledge generalize beyond U.S. Facebook users. The uniqueness of the context in which our study was conducted, in terms of BiH's history and intergroup relations, makes it an important test case for assessing the effect of social media usage on intergroup relations, but also requires that the findings be interpreted with the characteristics of the environment in mind. For now, our findings should be interpreted as being limited to a post-conflict area that is characterized by varying levels of spatial group segregation and no language comprehension barrier between

[‡‡] We conduct the same test with a continuous variable – Table S21 – and the insight continues to hold.

the members of main ethnic groups, as well as during an emotionally and politically charged period. Moreover, those actively using Facebook to propagate hatred and extremist views, of whom there are many, would possibly not show interest in participating in a study led by a U.S.-based university, and may thus not be fully captured within our sample. As such, our results do not speak to all the possible effects that Facebook could have even in BiH, and certainly not to all post-conflict societies. Isolating the ways in which the features of our research design shaped the direction of our main effect would be a valuable next step in advancing understanding of social media's impact on out-group attitudes.

Our study is, of course, not without limitations. The relatively small sample size reduces our power to better explore heterogeneity behind the observed effects. Replicating the study with a larger number of participants, both within BiH and in other countries, would be useful in understanding the circumstances in which social media is likely to worsen or ameliorate inter-ethnic attitudes. We also would encourage future research to compare the diversity of online versus offline networks, which would better shed light on the mechanisms underlying the relationship between the time spent on social media and group attitudes. In particular, additional measures of networks that go beyond the experiences of direct contact would allow for a finer-grained analysis of the differences in outcomes depending on the composition of one's network (although for now such studies on Facebook would likely require the cooperation of the company due to data access limitations for those outside of Facebook). Finally, as we deliberately set our study during a period of time around the Srebrenica genocide commemoration in which we expected emotions to be heightened and political discussions intensified, it remains to be seen how robust the effects we observed will be in a less emotionally charged period. Nevertheless, insufficient attention has been given to exploring the role that social media plays in enabling individuals to negotiate identity, preserve history, and cope with grief and societal trauma, a question to which we hope to contribute to answering in our future work.

Within the growing disappointments and valid concerns about the role of social media – many of which extend far beyond the scope of our paper – our findings do suggest that simply deactivating from social media is not a panacea to ethnic polarization, especially if the offline environment provides little to no opportunities for positive intergroup contact. Indeed, to answer the question of how time spent on social media affects users, scholars would do well to move away from social media determinism by paying attention to the contextual factors, alternative activities, and intergroup contact opportunities available to the individuals within the particular research context. How to continue doing so in a manner that can ultimately create a road map for strengthening of social capital through the tools of digital networks remains a crucial question in pushing this ever-important research agenda forward.

ACKNOWLEDGMENTS. We are grateful to all members of the NYU Center for Social Media and Politics, and participants of NYU Comparative Politics Workshop and CSMaP Election Seminar Series for their excellent comments and suggestions. We thank Cyrus Samii, Gwyneth McClendon, Chris Dawes, Matthew Gentzkow, Andrew Guess, and Kevin Munger for providing helpful feedback; Megan Brown for her excellent technical support; and Edima Zejnilovic for invaluable assistance on the ground.

1. JA Tucker, et al., Social media, political polarization, and political disinformation: A review of the scientific literature. (2018).
2. H Allcott, L Braghieri, S Eichmeyer, M Gentzkow, The welfare effects of social media. *Am. Econ. Rev.* **110**, 629–76 (2020).
3. N Persily, JA Tucker, eds., *Social Media and Democracy: The State of the Field and Prospects for Reform*. (Cambridge University Press), (2020).
4. S Iyengar, Y Lelkes, M Levendusky, N Malhotra, SJ Westwood, The origins and consequences of affective polarization in the united states. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2019).
5. A Urman, Context matters: political polarization on twitter from a comparative perspective. *Media, culture & society*, 0163443719876541 (2019).
6. J Poushter, C Bishop, H Chwe, Social media use continues to rise in developing countries but plateaus across developed ones. *Pew Res. Cent.* **22** (2018).
7. S Richey, J Zhu, Internet access does not improve political interest, efficacy, and knowledge for late adopters. *Polit. Commun.* **32**, 396–413 (2015).
8. F Campante, R Durante, F Sobbrio, Politics 2.0: The multifaceted effect of broadband internet on political participation. *J. Eur. Econ. Assoc.* **16**, 1094–1136 (2018).
9. CS Bailard, A field experiment on the internet's effect in an african election: Savvier citizens, disaffected voters, or both? *J. Commun.* **62**, 330–344 (2012).
10. A Nisser, Ph.D. thesis (2017).
11. SJ Westwood, et al., The tie that divides: Cross-national evidence of the primacy of partyism. *Eur. J. Polit. Res.* **57**, 333–354 (2018).
12. S Iyengar, SJ Westwood, Fear and loathing across party lines: New evidence on group polarization. *Am. J. Polit. Sci.* **59**, 690–707 (2015).
13. TF Pettigrew, LR Tropp, A meta-analytic test of intergroup contact theory. *J. personality social psychology* **90**, 751 (2006).
14. EL Paluck, SA Green, DP Green, The contact hypothesis re-evaluated. *Behav. Public Policy* **3**, 129–158 (2019).
15. M Hewstone, H Swart, Fifty-odd years of inter-group contact: From hypothesis to integrated theory. *Br. J. Soc. Psychol.* **50**, 374–386 (2011).
16. M Wojcieszak, BR Warner, Can interparty contact reduce affective polarization? a systematic test of different forms of intergroup contact. *Polit. Commun.*, 1–23 (2020).
17. TD Baruah, Effectiveness of social media as a tool of communication and its potential for technology enabled connections: A micro-level study. *Int. J. Sci. Res. Publ.* **2**, 1–10 (2012).
18. Z Papacharissi, The virtual sphere: The internet as a public sphere. *New media & society* **4**, 9–27 (2002).
19. R Plant, Online communities. *Technol. society* **26**, 51–65 (2004).
20. H Rainie, B Wellman, *Networked: The new social operating system*. (Mit Press Cambridge, MA) Vol. 419, (2012).
21. H Baytiyeh, Social media's role in peacebuilding and post-conflict recovery. *Peace Rev.* **31**, 74–82 (2019).
22. CT Carr, Social media and intergroup communication in *Oxford Research Encyclopedia of Communication*. (2017).
23. MM Metzger, R Bonneau, J Nagler, JA Tucker, Tweeting identity? ukrainian, russian, and# euromaidan. *J. Comp. Econ.* **44**, 16–40 (2016).
24. M Granovetter, The strength of weak ties: A network theory revisited. *Sociol. theory*, 201–233 (1983).
25. RD Putnam, , et al., *Bowling alone: The collapse and revival of American community*. (Simon and schuster), (2000).
26. D Williams, On and off the'net: Scales for social capital in an online era. *J. computer-mediated communication* **11**, 593–628 (2006).
27. Y Amichai-Hamburger, T Hayat, Social networking. *The international encyclopedia media effects*, 1–12 (2017).
28. GW Allport, K Clark, T Pettigrew, The nature of prejudice. (1954).
29. D Williams, The impact of time online: Social capital and cyberbalkanization. *CyberPsychology & behavior* **10**, 398–406 (2007).
30. ML Antheunis, MM Vanden Abeele, S Kanters, The impact of facebook use on micro-level social capital: A synthesis. *Societies* **5**, 399–419 (2015).
31. KM Dowley, BD Silver, Social capital, ethnicity and support for democracy in the post-communist states. *Eur. Stud.* **54**, 505–527 (2002).
32. P M. Pickering, Generating social capital for bridging ethnic divisions in the balkans: Case studies of two bosniak cities. *Ethn. Racial Stud.* **29**, 79–103 (2006).
33. K Faucher, *Social Capital Online*. (University of Westminster Press), (2018).
34. M Hindman, *The myth of digital democracy*. (Princeton University Press), (2008).
35. H Farrell, The consequences of the internet for politics. *Annu. review political science* **15** (2012).
36. I Tellidis, S Kappler, Information and communication technologies in peacebuilding: Implications, opportunities and challenges. *Coop. Confl.* **51**, 75–93 (2016).
37. CK Morewedge, Negativity bias in attribution of external agency. *J. Exp. Psychol. Gen.* **138**, 535 (2009).
38. P Rozin, EB Royzman, Negativity bias, negativity dominance, and contagion. *Pers. social psychology review* **5**, 296–320 (2001).
39. JT Cacioppo, WL Gardner, Emotion. *Annu. review psychology* **50**, 191–214 (1999).
40. JE Settle, *Frenemies: How social media polarizes America*. (Cambridge University Press), (2018).
41. KN Hampton, Social media and change in psychological distress over time: The role of social causation. *J. Comput. Commun.* **24**, 205–222 (2019).
42. A Alter, *Irresistible: The rise of addictive technology and the business of keeping us hooked*. (Penguin), (2017).
43. A Błachnio, A Przepiórka, I Pantic, Internet use, facebook intrusion, and depression: Results of a cross-sectional study. *Eur. Psychiatry* **30**, 681–684 (2015).
44. International Court of Justice, Application of the convention on the prevention and punishment

of the crime of genocide (bosnia and herzegovina v. serbia and montenegro) (2007) https://www.icj-cij.org/en/case/91/judgments.
45. S GlobalStats, Social media stats bosnia and herzegovina. *StatCounter Glob.* (2020).
46. W Lin, , et al., Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *The Annals Appl. Stat.* **7**, 295–318 (2013).
47. GW Imbens, DB Rubin, *Causal inference in statistics, social, and biomedical sciences*. (Cambridge University Press), (2015).
48. Y Benjamini, AM Krieger, D Yekutieli, Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507 (2006).
49. L Boxell, M Gentzkow, JM Shapiro, Cross-country trends in affective polarization, (National Bureau of Economic Research), Technical report (2020).
50. S DellaVigna, E La Ferrara, Economic and social impacts of the media in *Handbook of media economics*. (Elsevier) Vol. 1, pp. 723–768 (2015).
51. G Dahl, S DellaVigna, Does movie violence increase violent crime? *The Q. J. Econ.* **124**, 677–734 (2009).
52. M Gentzkow, JM Shapiro, Preschool television viewing and adolescent test scores: Historical evidence from the coleman study. *The Q. J. Econ.* **123**, 279–323 (2008).
53. B Rimé, C Finkenauer, O Luminet, E Zech, P Philippot, Social sharing of emotion: New evidence and new questions. *Eur. review social psychology* **9**, 145–189 (1998).
54. JR Vollhardt, Collective victimization. *Oxf. handbook intergroup conflict*, 136–157 (2012).
55. A Guess, B Nyhan, B Lyons, J Reifler, Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think. *Knight Foundation White Pap.* (2018).
56. M Gentzkow, JM Shapiro, Ideological segregation online and offline. *The Q. J. Econ.* **126**, 1799–1839 (2011).
57. O Christ, et al., Contextual effect of positive intergroup contact on outgroup prejudice. *Proc. Natl. Acad. Sci.* **111**, 3996–4000 (2014).
58. DA Scheufele, BW Hardy, D Brossard, IS Waismel-Manor, E Nisbet, Democracy based on difference: Examining the links between structural heterogeneity, heterogeneity of discussion networks, and democratic citizenship. *J. Commun.* **56**, 728–753 (2006).
59. A Gelman, J Carlin, Beyond power calculations: Assessing type s (sign) and type m (magnitude) errors. *Perspectives on Psychol. Sci.* **9**, 641–651 (2014).
60. M Bailey, et al., The determinants of social connectedness in europe in *International Conference on Social Informatics*. (Springer), pp. 1–14 (2020).
61. AS Hodzic, Sanela, Media sustainability index, (IREX), Technical report (2018).