nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01655-0

Reducing political polarization in the United States with a mobile chat platform

Received: 26 September 2022

Accepted: 14 June 2023

Published online: 21 August 2023

Check for updates

Aidan Combs[1,8], Graham Tierney[2,8], Brian Guay[3,4], Friedolin Merhout[5], Christopher A. Bail[1,6], D. Sunshine Hillygus[6,7] & Alexander Volfovsky[2]

Do anonymous online conversations between people with different political views exacerbate or mitigate partisan polarization? We created a mobile chat platform to study the impact of such discussions. Our study recruited Republicans and Democrats in the United States to complete a survey about their political views. We later randomized them into treatment conditions where they were offered financial incentives to use our platform to discuss a contentious policy issue with an opposing partisan. We found that people who engage in anonymous cross-party conversations about political topics exhibit substantial decreases in polarization compared with a placebo group that wrote an essay using the same conversation prompts. Moreover, these depolarizing effects were correlated with the civility of dialogue between study participants. Our findings demonstrate the potential for well-designed social media platforms to mitigate political polarization and underscore the need for a flexible platform for scientific research on social media.

Political polarization is one of the most pressing social problems of our era[1–8]. Though scholars were once optimistic that the internet could help bridge partisan divides by allowing people to connect with broader communities, many now worry that social media platforms have increased ideological segregation and incivility instead[9–13]. Understanding whether and under what conditions online communication across party lines can shape political divisions is critical to addressing the challenges facing democracy.

Past research provides mixed evidence about the role of cross-party interactions in political polarization. Some studies suggest that these interactions can exacerbate polarization and incivility[14,15]. For instance, recent work concludes that Facebook use may increase political polarization[13], and exposure to ideologically uncongenial information can push partisans further apart[16]. Yet other research suggests that people moderate their views when they engage with those with different perspectives because they come to recognize the value of alternative viewpoints[17–23].

Particularly little is known about the impact of cross-party conversations that occur in online settings[24]. Previous research indicates that online communications are often less civil because people feel less encumbered by the social norms that guide physical interaction[25,26], which can lead them to more easily dehumanize others who disagree with them[27]. This may be especially the case if they are partly or completely anonymous to each other[28–30]. However, there is also evidence that anonymity can encourage people to focus on the content of conversations rather than the identity of the people with whom they engage[31–34]. Anonymity may also allow people to explore alternative viewpoints honestly without fear of social repercussions[15,35,36]. Research is therefore needed to determine whether anonymous cross-party conversations in online settings will exacerbate or mitigate political polarization.

Unfortunately, studying the causal effects of online cross-party conversations presents considerable methodological challenges. Observational analyses of cross-party deliberations on social media

[1]Department of Sociology, Duke University, Durham, NC, USA. [2]Department of Statistical Science, Duke University, Durham, NC, USA. [3]Department of Political Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Sociology, University of Copenhagen, Copenhagen, Denmark. [6]Sanford School of Public Policy, Duke University, Durham, NC, USA. [7]Department of Political Science, Duke University, Durham, NC, USA. [8]These authors contributed equally: Aidan Combs, Graham Tierney. e-mail: alexander.volfovsky@duke.edu

platforms are poorly suited to identifying such effects because the processes that lead people into such interactions are not random[16,37–39]. Moreover, platforms such as Facebook or Twitter are typically unwilling to randomize their users into the experiments necessary to test hypotheses about the potential effects because of corporate priorities to protect user privacy and ensure consistent user experience[40–43].

To address these issues, we developed our own mobile chat platform to conduct a field experiment testing the impact of anonymous cross-party conversations on controversial topics. Creating our own platform allowed us to customize and randomly assign features of the user experience, while holding constant other features that might impact online behaviour. As described below, we paired Republicans and Democrats to complete a sustained, text-based conversation about a political topic using our platform. The participants received varied information about the political preferences of their chat partners, with whom they discussed immigration or gun control over a period of several days.

We designed our platform to emulate the look and feel of a social media app. The platform, called DiscussIt, enables asynchronous, text-based messages similar to WhatsApp or direct messaging on Facebook or Twitter. The participants downloaded the app to their mobile devices from the Google Play or Apple App Store. The platform featured visual images created by a graphic designer and a staff-supported user support and content moderation team. The participants could like each other's messages (though we did not show them whether their messages had been liked by their partner), enable notifications and block or report their chat partner. The platform also allowed embedded survey questions and the collection of behavioural data. Combined with ostensibly unrelated pre and post surveys outside of the platform, the experimental design allowed us to rigorously evaluate the causal impact of anonymous cross-party conversations on multiple dimensions of political polarization.

Results

Procedure

In early February 2020, we hired the survey firm YouGov to recruit self-identified Democrats and Republicans from their survey panel to participate in our field experiment. The participants started with a survey about their political views, which included multiple questions measuring both issue polarization (that is, ideological polarization) and affective polarization (the gap between individuals' positive feelings towards their own political party and negative feelings towards the opposing party)[7]. The issue polarization measure was specific to the particular topic assigned for discussion—either immigration or gun control. Similar to previous research[44], our outcome of interest is a global index, constructed from all 21 measures of polarization (Cronbach's $\alpha = 0.72$), as well as sub-indices that report our results separately for issue-based and affective polarization. These indices improve measurement precision[45], and breaking them into sub-indices allows us to examine the effects of different types of polarization separately[46,47]. Each of our depolarization indices is coded such that more positive values indicate people expressing less polarized views. We provide additional details about question wording and scale construction in the Supplementary Information as well as multiple robustness checks.

After the participants completed the survey, we randomly assigned them to treatment and control conditions and sent the treatment group a seemingly unrelated invitation to download and test a mobile app for a new social media platform called DiscussIt for financial compensation. Our seemingly unrelated design is an advance over previous work in that it both obscures the political nature of the experimental treatment and guards against demand effects. The invitation informed the participants that DiscussIt is a social media platform where people anonymously discuss various topics but instructed them not to disclose their name or personal information in order to allow conversations to "develop freely". There was no mention of politics in the recruitment dialogue. Participants who downloaded the app ($n = 1{,}201$) were assigned an 'invite' code that—unbeknownst to them—automatically paired them with an opposing partisan. We provide details of the recruitment process and challenges, including a lower-than-expected yield of participants, and further evaluations of the resulting sample in the Supplementary Information. Figure 1 shows the on-boarding app screens welcoming the participants, which instructed them to complete 14 thoughtful replies with another DiscussIt user over the course of one week to receive the financial incentive. They were then randomly assigned to one of the two discussion topic areas—immigration or gun control—and given a gender-neutral pseudonym. After they were matched with an opposing partisan, the participants advanced to a chat screen, where they received a prompt to begin the conversation about the designated policy topic.

The participants were further randomized into one of three labelling conditions in the treatment group: the discussion partner was (1) correctly identified as an opposing partisan, (2) not labelled with a party or (3) mislabelled to have the same party identification as the participant. These sub-conditions were meant to provide further insight about how information on partisan identity might shape anonymous conversations about politics[8,22,46,48]. Participants in the control group received a placebo prompt asking them to write an essay on immigration or gun control in response to the same prompts provided at the end of onboarding in the app. The aim of this baseline condition is to ensure all individuals in the study (both treated and not treated) have given roughly equivalent thought to the specific policy topic[49]. See the Supplementary Information for analyses using a separate control condition where participants were not asked to engage in any activity about the policy topic.

Several days after participants in the treatment condition completed their chats, all participants received another invitation to an ostensibly unrelated survey about health. This survey included the same measures used in the pre-treatment survey, thereby enabling within-subject assessment of the impact of our intervention, but began with a set of distractor questions designed to mask the purpose of our study and discourage demand effects based on the participants' interpretation of the aims of our study. Finally, our app also collected the full text of all conversations, allowing further analysis of the potential mechanisms that shape anonymous cross-party interaction on social media.

Analyses

Our analysis estimated the change between our pre-treatment and post-treatment depolarization index via a two-stage least-squares model designed to assess the complier average causal effect (CACE) of our intervention[50]. Comparisons between compliant and non-compliant participants are reported in Supplementary Information section 4.3, and intent-to-treat results are presented in Supplementary Information section 4.4. There are no differences in the statistical significance of effects observed between the CACE and intent-to-treat models. As preregistered, compliance was defined as installing our study's app and completing at least ten exchanges with a member of the opposing party. As Fig. 2 shows, we found that participants in our treatment condition exhibited sizable increases in our depolarization index relative to the placebo condition even after relatively short conversations on our platform—equivalent to 0.22 standard deviations ($t = 2.44$; $n = 1{,}419$; two-tailed $P = 0.016$; 95% confidence interval (CI), 0.0426 to 0.411). The increase was roughly equal on both affect and issue sub-indices, though the increase is statistically significant only for the issue index.

Figure 3 reports the treatment effects by labelling condition (that is, whether or how the partisanship of the participant's conversation partner was labelled). Participants in the correct-labels condition exhibited a 0.25-standard-deviation increase ($t = 2.50$; $n = 1{,}419$; two-tailed $P = 0.012$; 95% CI, 0.055 to 0.457), and those in the incorrect-labels

https://doi.org/10.1038/s41562-023-01655-0

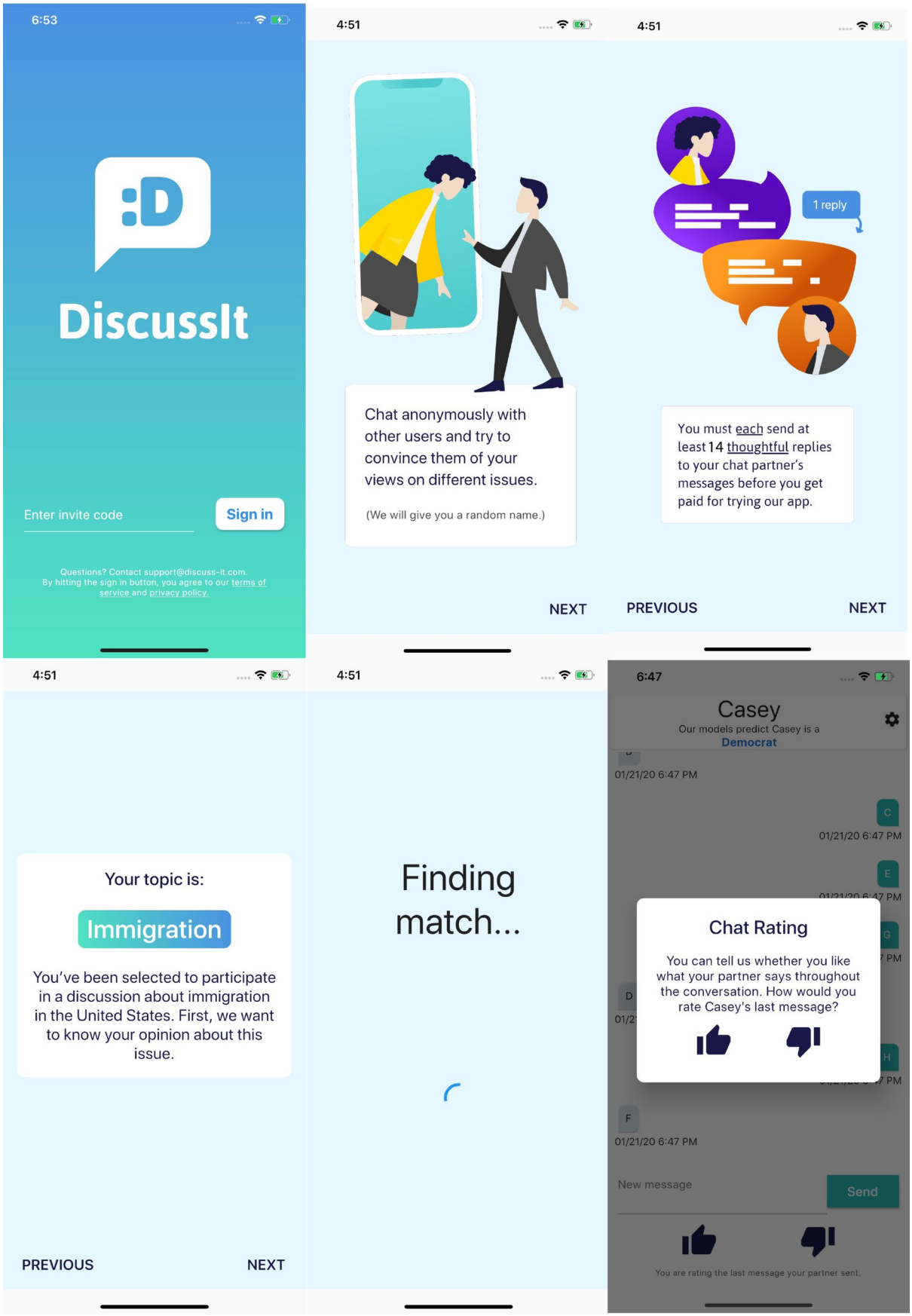

Fig. 1 | Onboarding images from the social media platform created for this study. After downloading the app, the participants logged in and were guided through several onboarding screens. They were then shown their randomly assigned discussion prompt about either immigration or gun control and matched with a partner from the opposing political party. Whether their partner's party affiliation was displayed correctly, incorrectly or not at all was randomized at the time of matching. After matching, the participants entered the chat interface and could begin their conversation. Images adapted from www.humaaans.com under Creative Commons CC0.

https://doi.org/10.1038/s41562-023-01655-0

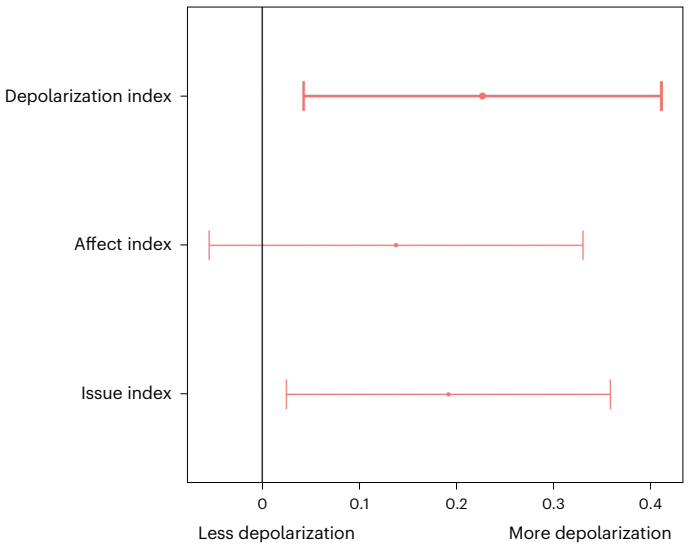

Fig. 2 | Effect of cross-party interaction on an anonymous chat platform on political polarization. The research participants (n = 715 Democrats, n = 704 Republicans) were randomized to either a treatment condition, where they were invited to test an app for a new social media platform that (unbeknownst to them) paired them with an opposing partisan to discuss immigration or gun control policy, or a placebo condition. In the placebo condition, participants wrote an essay using the same conversation prompts as the treatment condition. Outcomes were measured on a depolarization index that is coded to be positive if participants expressed less polarized attitudes. The indices are standardized to variance 1 for each analysis, so the effect on the depolarization index does not necessarily lie between the effects on the component indices. The plot shows the point estimates and 95% CIs for the change in depolarization among those in the treatment relative to those in the placebo condition (change in depolarization in treatment minus change in placebo condition) for the global index (in bold) and the issue and affect sub-indices. We observed a significant and positive effect of cross-party interaction. This finding indicates that using the study's anonymous chat platform to discuss a political issue with an opposing partisan depolarized participants by approximately 0.22 standard deviations on the depolarization index (t = 2.44; n = 1,419; two-tailed P = 0.016; 95% CI, 0.0426 to 0.411). The effect is positive on both the issue and affect sub-indices but is statistically significant only for the issue index. For the affect and issue indices, respectively, CACE = 0.138 and 0.191; two-tailed P = 0.160 and 0.024; t statistics, 1.41 and 2.25 on 1,417 and 1,411 degrees of freedom; 95% CIs, −0.055 to 0.330 and 0.0249 to 0.359. Standard errors are clustered at the conversation level. The CIs are centred at the point estimates. See Supplementary Information section 3 for the full model details.

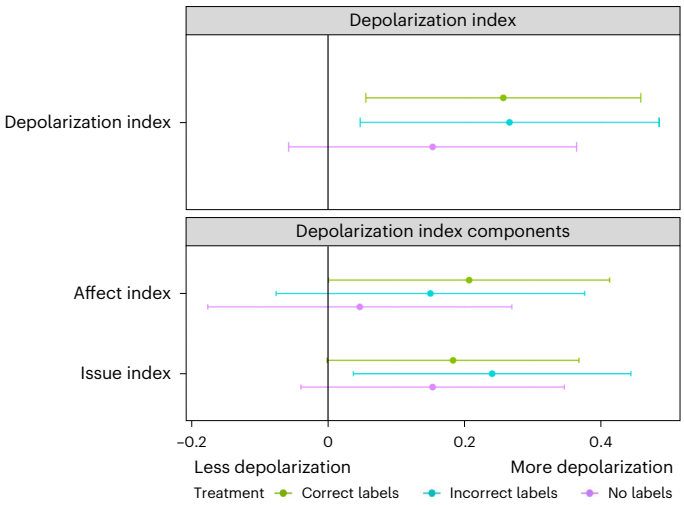

Fig. 3 | Effect of cross-party interaction on an anonymous chat platform on political polarization according to different identity cues. The treatment effects by different labelling conditions relative to the placebo condition are shown. The green, teal and purple bars indicate the treatment effects for those whose discussion partner was accurately labelled as an opposing partisan, was mislabelled as a co-partisan and had their party affiliation unlabelled, respectively. The top panel shows the effects on the depolarization index, and the bottom panel shows the effects on the issue and affect sub-indices. We observed large and significant effects for the correctly and incorrectly labelled conditions on the depolarization index (CACE = 0.253 and 0.262, respectively; two-tailed P = 0.012 and 0.017; t statistics, 2.56 and 2.39 on 1,415 degrees of freedom; 95% CIs, 0.055 to 0.457 and 0.047 to 0.484) and insignificant effects for the unlabelled condition (CACE = 0.0151; two-tailed P = 0.15; t statistic, 1.42 on 1,415 degrees of freedom; 95% CI, −0.058 to 0.363). The correct-labels results for the affect and issue indices, respectively, were CACE = 0.206 and 0.183; two-tailed P = 0.049 and 0.051; t statistics, 1.97 and 1.95 on 1,415 and 1,409 degrees of freedom; 95% CIs, 0.001 to 0.412 and −0.002 to 0.367. The incorrect-labels results for the affect and issue indices, respectively, were CACE = 0.150 and 0.240; two-tailed P =0.193 and 0.020; t statistics, 1.30 and 2.32 on 1,415 and 1,409 degrees of freedom; 95% CIs, −0.0759 to 0.375 and 0.037 to 0.443. The no-labels results for the affect and issue indices, respectively, were CACE = 0.046 and 0.153; two-tailed P = 0.682 and 0.120; t statistics, 0.41 and 1.56 on 1,415 and 1,409 degrees of freedom; 95% CIs, −0.176 to 0.269 and −0.040 to 0.346. Standard errors are clustered at the conversation level. The CIs are centred at the point estimates. The indices are standardized to variance 1 for each analysis, so the effect on the depolarization index does not necessarily lie between the effects on the component indices. See Supplementary Information section 3 for the full model details.

condition showed a 0.26-standard-deviation increase (t = 2.34; n = 1,419; two-tailed P = 0.017; 95% CI, 0.047 to 0.484) in our depolarization index. Participants in the sub-condition where political parties were unlabelled exhibited a 0.15-standard-deviation treatment effect, which was not statistically significant (t = 1.43; n = 1,419; two-tailed P = 0.15; 95% CI, −0.058 to 0.363); this result reflects treatment effects of nearly zero (0.05 standard deviations; t = 0.41; n = 1,419; two-tailed P = 0.68; 95% CI, −0.176 to 0.269) for the unlabelled condition on the affect index. The point estimates of treatment effects are much more similar for the issue index. Figure 4 reports the treatment effects by the political party of the participant. The treatment effect was approximately 0.38 standard deviations for Republicans (t = 2.68; n = 704; two-tailed P = 0.007; 95% CI, 0.101 to 0.650) and 0.09 for Democrats (not statistically significant; t = 0.697; n = 715; two-tailed P = 0.48; 95% CI, −0.156 to 0.328). In the Supplementary Information, we show that Republicans were less polarized before treatment than Democrats, suggesting that Republicans did not simply have more room to moderate than Democrats. We are unable to reject the null hypothesis that the treatment effect was the same for Republicans and Democrats (treatment effect difference, 0.29; z = 1.55; two-tailed P = 0.12; 95% CI, −0.0768 to 0.656). Results are also shown for the individual components of the polarization index, with treatment effect point estimates of 0.4 standard deviations for Republicans and 0.01 for Democrats for the issue index and 0.18 for Republicans and 0.08 for Democrats for the affect index.

Finally, we analysed the civility of the messages exchanged using natural language processing techniques[51], an approach that has been used in previous research examining interpersonal exchanges about political disagreements[52]. Figure 5 describes the 'civility index' created via this analysis for participants' chat partners over time. As this figure shows, people who experienced significant increases in our depolarization index tended to have conversation partners who used more civil language—particularly during the beginning of the conversation. In the Supplementary Information, we present models showing that the relationship between the chat partner's civility and depolarization is statistically significant. We note that civility was not incorporated into our randomization design and therefore cannot be considered an unambiguous causal factor.

https://doi.org/10.1038/s41562-023-01655-0

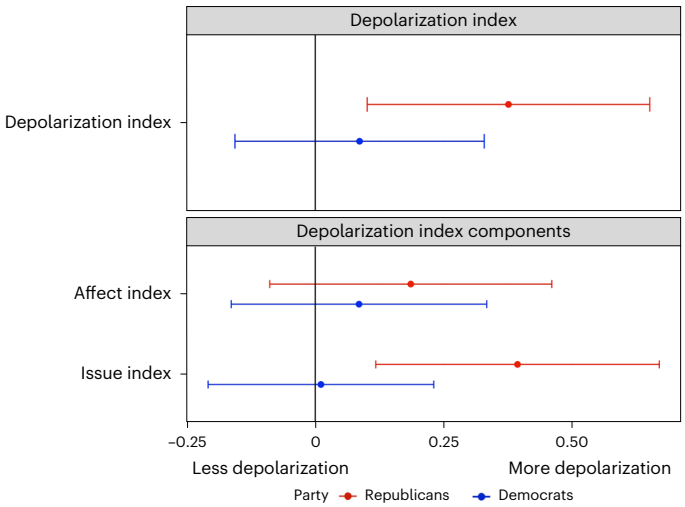

Fig. 4 | Effect of cross-party interaction on an anonymous chat platform on political polarization, by party. The effects of our intervention are shown for Republicans (CACE = 0.375; $t = 2.68$; $n = 704$; two-tailed $P = 0.007$; 95% CI, 0.101 to 0.650) and Democrats (CACE = 0.086; $t = 0.697$; $n = 715$; two-tailed $P = 0.48$; 95% CI, −0.156 to 0.328). We cannot reject the null hypothesis that the two treatment effects are equal (treatment effect difference, 0.289; $z = 1.55$; two-tailed $P = 0.12$; 95% CI, −0.0768 to 0.656). The affect index results for Republicans and Democrats, respectively, were CACE = 0.186 and 0.085; two-tailed $P = 0.18$ and 0.50; t statistics, 1.32 and 0.67 on 702 and 713 degrees of freedom; 95% CIs, −0.0884 to 0.460 and −0.163 to 0.333. The issue index results for Republicans and Democrats, respectively, were CACE = 0.393 and 0.011; two-tailed $P = 0.005$ and 0.92; t statistics, 2.80 and 0.10 on 700 and 709 degrees of freedom; 95% CIs, 0.118 to 0.669 and −0.208 to 0.230. The CIs are centred at the point estimates. Traditional standard errors are reported because members of the same party did not interact. The indices are standardized to variance 1 for each analysis, so the effect on the depolarization index does not necessarily lie between the effects on the component indices. See Supplementary Information section 3 for the full model specification details.

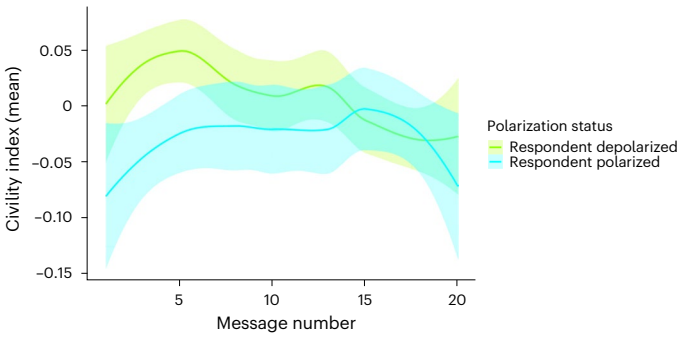

Fig. 5 | Civility by treatment outcome (over time). Each chat produced on our platform was analysed using natural language processing software to identify the frequency of civil exchanges. This figure plots the resultant civility index for each participant's chat partner's messages according to their polarization status. Depolarized participants are those whose depolarization index was more than one standard deviation above the mean; polarized participants' depolarization index was more than one standard deviation below the mean. Participants who depolarized tended to have chat partners who used more civil language, particularly at the beginning of the conversation. The time series data were smoothed with a LOESS function.

Discussion

Our findings show that anonymous online cross-party conversations can help depolarize the public. This finding is particularly noteworthy since there are currently so few examples of interventions that successfully reduce political tribalism online[7,53]. Importantly, this depolarization occurred without explicit appeals for deliberation, empathy or cooperation by the intervention—more closely reflecting the user experiences on social media.

The final sample sizes in our experiment allow for comparisons of each sub-condition to the placebo condition but do not enable precise comparisons across sub-conditions without strong parametric assumptions. Nonetheless, the observed patterns offer some suggestions as to the potential mechanisms underlying our findings. The smallest treatment effects are found among those whose partisan identifications were unlabelled. This pattern suggests that participants in conversations with explicit partisan labels may more easily draw connections between the conversation and existing partisan stereotypes, experiences and attitudes—a necessary precondition for attitudinal change. This interpretation is also reflected in the fact that the point estimate of the unlabelled condition's effect on the issue index is larger and more comparable to the other treatment conditions than its effect on the affect index, although those results are both not statistically significant. This finding is consistent with recent research revealing that polarization is fuelled by wildly overestimated partisan stereotypes[54–59], so that a conversation with a member of the other party that contradicts prevailing stereotypes—revealing the actual extent of heterogeneity in partisan views—should help depolarize[21,23,60–62]. It is telling that this is the case whether a participant is told they are talking to someone from their outgroup or ingroup—consistent with recent research finding that ingroup social pressures contribute to polarization[63]. Analyses of the content of the conversations with respect to civility further reinforce this interpretation.

We also examined heterogeneity in the treatment effects by partisanship. A growing number of studies indicate that political polarization in the United States has evolved in an asymmetric manner, driven primarily by Republicans. For example, Republican elected officials have increasingly taken more extreme positions in legislative votes than their counterparts in the Democratic party over the past 40 years[64]. Other studies indicate that exposing Republicans to Democrats can make them more polarized, not less[16,33,65]. Yet these studies exposed people to high-profile elites or studied the effect of exposure in larger group settings or with less sustained interactions than those used here. It may be that anonymous dyadic conversation between non-elites is a particularly important counterweight to asymmetric polarization by providing Republicans with the space to encounter views and stereotypes distinct from elite rhetoric and conservative media[66]. In other words, people might find it easier to find common ground with a regular person than a political elite[23]—about whom they have strong stereotypes generated by partisan media. Indeed, our results are consistent with work by Balietti and colleagues showing that people with conservative views depolarize more than those with liberal views when exposed to cross-attitudinal arguments written by peers rather than elites[22].

Our research has some notable limitations. Though our participants were recruited from a high-quality online survey panel, they are not representative of the general population or of social media users (Supplementary Information section 2.3). It is possible that the people who participated in our experiment—that is, those who were willing to download and use an unknown app to converse with strangers—respond differently to those conversations than a typical social media user.

Additionally, the cross-party exchanges on the DiscussIt platform are not a perfect analogue for discussions on social media. Although the platform mimicked the look and feel of a social media app in many ways, key differences remain. For example, anonymous conversations might unfold quite differently in a non-dyadic setting, where larger numbers of users interacting may generate peer influence dynamics that are quite different and where reputational concerns are more salient. In an open setting, members of one party may feel compelled

to team up on the other because of established social norms or the possibility of receiving affirmation via 'likes' from their ingroup. In addition, we instructed our participants to engage in a substantial number of back-and-forth replies to each other. Cross-party interactions on established platforms are often much shorter, and it may be that sustained conversation is necessary for depolarization. Finally, we asked the participants to engage in a focused discussion about a particular issue. Our intervention may thus more closely resemble a platform where discussions might be organized around a particular topic.

Nevertheless, our research demonstrates the promise of creating a social media platform for scientific research. We show that online conversations on a platform with the look and feel of contemporary social media can depolarize participants under the particular set of conditions we created. Specifically, we found that conversation between partisans that is dyadic, anonymous, thoughtful and focused on political issues depolarizes participants. An important and fruitful path for future work will be to further investigate the effects of conversation under other conditions reflective of contemporary platforms. Tools like the one that we created would enable that work.

Tools like ours not only could be used to conduct high-quality field experiments that examine many other design features but also could avoid the many challenges of collaborating with social media platforms to conduct research on the increasingly urgent topic of political polarization. Perhaps most importantly, this research paradigm may inspire scientists, entrepreneurs or existing social media companies to explore entirely new design features. Most of the dominant platforms evolved in a chaotic manner, where interventions are tested to address emerging threats. In contrast, a new research agenda focused on scientifically testing the impact of social media design on polarization has the potential to test and develop a fuller range of design features that might incentivize more positive behaviour.

Methods

Our research was approved by the Institutional Review Board at Duke University (protocol no. 2020-0326) and preregistered on the Open Science Framework on 20 January 2020 (https://osf.io/g97z5/). Informed consent was collected from all participants, and all participants were compensated via the survey firm YouGov. Recruiting participants to download an app for an unrecognized mobile chat platform is difficult. We describe these challenges in Supplementary Information section 2.

The participants were randomized into the treatment or placebo condition, and those assigned to treatment were then further randomized into one of the three display conditions. Sample sizes were chosen to have 80% power to detect effects of at least 0.25 standard deviations. A total of 1,419 participants (640 men, 779 women; average age, 52.3 years) were included in our analyses. Data collection was done blind to treatment condition, and the ostensibly unrelated surveys were administered by YouGov. As is perhaps inevitable given the complexity of the field experiment, some of these implementation issues necessitated deviation from the preregistration. Specifically, the final sample sizes for each sub-condition sufficiently power a comparison to the baseline but offer limited precision for comparisons to one another. The deviation occurred because individuals in the control condition were not initially screened for their willingness to download an app, which was an inclusion criterion for our analysis. Due to the ostensibly unrelated nature of all of our surveys and since those in the control condition were not exposed to any treatment, those in the control condition required re-surveying to measure their willingness to participate. Lastly, as all screened respondents completed a pre-treatment survey outside the app before being issued an invitation to download the app, we were left with a smaller-than-anticipated sample size when we had a lower yield of survey respondents following through on the app download.

All analyses in this manuscript were performed in R 4.1.1 on a 2020 M1 MacBook Pro computer. The CACEs reported in the main text were computed using two-stage least squares. All regression-adjusted models include main effects and cluster robust standard errors. Detailed model specifications, as well as specification of the analysis for intent-to-treat analyses (using linear regression) and heterogeneous treatment effect analysis (using Bayesian additive regression trees), are provided in the Supplementary Information.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Anonymized replication data are publicly available from the authors at this link: https://doi.org/10.7910/DVN/LTVEHJ.

Code availability

Replication code for the main results in the manuscript is publicly available at this link: https://doi.org/10.7910/DVN/LTVEHJ.

References

1. Voelkel, J., Stagnaro, M., Chu, J., Pink, S. & Mernyk, J. *Megastudy Identifying Successful Interventions to Strengthen Americans' Democratic Attitudes* (Institute for Policy Research Working Papers, 2022).
2. Baldassarri, D. & Bearman, P. Dynamics of political polarization. *Am. Sociol. Rev.* **72**, 784–811 (2007).
3. Boxell, L., Gentzkow, M. & Shapiro, J. *Cross-Country Trends in Affective Polarization* Working Paper No. 26669 (National Bureau of Economic Research, 2020).
4. DellaPosta, D., Shi, Y. & Macy, M. Why do liberals drink lattes? *Am. J. Sociol.* **120**, 1473–1511 (2015).
5. Finkel, E. J. et al. Political sectarianism in America. *Science* **370**, 533–536 (2020).
6. Fiorina, M. P. & Abrams, S. J. Political polarization in the American public. *Annu. Rev. Polit. Sci.* **11**, 563–588 (2008).
7. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. The origins and consequences of affective polarization in the United States. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2019).
8. Mason, L. *Uncivil Agreement* (Univ. Chicago Press, 2018).
9. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
10. Barberá, P. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Polit. Anal.* **23**, 76–91 (2015).
11. Sunstein, C. R. *Republic.com* (Princeton Univ. Press, 2002).
12. Levy, R. Social media, news consumption, and polarization: evidence from a field experiment. *Am. Econ. Rev.* **111**, 831–70 (2021).
13. Settle, J. E. *Frenemies: How Social Media Polarizes America* (Cambridge Univ. Press, 2018).
14. Papacharissi, Z. Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media Soc.* https://doi.org/10.1177/1461444804041444 (2004).
15. Price, V. in *Online Deliberation: Design, Research, and Practice* (eds Davies, T. & Gangadharan, S. P.) 37–58 (Univ. Chicago Press, 2009).
16. Bail, C. et al. Exposure to opposing views on social media can increase political polarization. *Proc. Natl Acad. Sci. USA* **115**, 9216–9221 (2018).
17. Fishkin, J. S. & Luskin, R. C. Experimenting with a democratic ideal: deliberative polling and public opinion. *Acta Polit.* **40**, 284–298 (2005).
18. Mutz, D. *Hearing the Other Side: Deliberative versus Participatory Democracy* (Cambridge Univ. Press, 2006).

19. Zhang, K. Encountering dissimilar views in deliberation: political knowledge, attitude strength, and opinion change. *Polit. Psychol.* **40**, 315–333 (2019).
20. Broockman, D. & Kalla, J. Durably reducing transphobia: a field experiment on door-to-door canvassing. *Science* **352**, 220–224 (2016).
21. Fishkin, J. S., Siu, A., Diamon, L. & Bradburn, N. Is deliberation an antidote to extreme partisan polarization? Reflections on 'America in One Room'. *Am. Polit. Sci. Rev.* **115**, 1464–1481 (2021).
22. Balietti, S., Getoor, L., Goldstein, D. G. & Watts, D. J. Reducing opinion polarization: effects of exposure to similar people with differing political views. *Proc. Natl Acad. Sci. USA* **118**, e2112552118 (2021).
23. Levendusky, M. S. & Stecula, D. A. *We Need to Talk* (Cambridge Univ. Press, 2021); https://www.cambridge.org/core/product/identifier/9781009042192/type/element
24. Santoro, E. & Broockman, D. E. The promise and pitfalls of cross-partisan conversations for reducing affective polarization: evidence from randomized experiments. *Sci. Adv.* **8**, eabn5515 (2022).
25. Cheng, J., Danescu-Niculescu-Mizil, C. & Leskovec, J. How community feedback shapes user behavior. In *Proceedings of the International AAAI Conference on Web and Social Media*, *Vol. 8* 41–50 (2014).
26. Kiesler, S., Siegel, J. & McGuire, T. W. Social psychological aspects of computer-mediated communication. *Am. Psychol.* **39**, 1123–1134 (1984).
27. Schroeder, J., Kardas, M. & Epley, N. The humanizing voice: speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychol. Sci.* **28**, 1745–1762 (2017).
28. Lowry, P. B., Zhang, J., Wang, C. & Siponen, M. Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Inf. Syst. Res.* **27**, 962–986 (2016).
29. Lapidot-Lefler, N. & Barak, A. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Comput. Hum. Behav.* **28**, 434–443 (2012).
30. Suler, J. The online disinhibition effect. *Cyberpsychol. Behav.* **7**, 321–326 (2004).
31. Berg, J. The impact of anonymity and issue controversiality on the quality of online discussion. *J. Inf. Technol. Polit.* **13**, 37–51 (2016).
32. De Choudhury, M. & De, S. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. In *Proc. 8th International AAAI Conference on Web and Social Media, Vol. 8* 41–50 (2014).
33. Guilbeault, D., Becker, J. & Centola, D. Social learning and partisan bias in the interpretation of climate trends. *Proc. Natl Acad. Sci. USA* **115**, 9714–9719 (2018).
34. Strandberg, K. & Berg, J. Impact of temporality and identifiability in online deliberations on discussion quality: an experimental study. *Javnost* **22**, 164–180 (2015).
35. Mansbridge, J. J. *Beyond Adversary Democracy* (Univ. Chicago Press, 1983).
36. Sanders, L. M. Against deliberation. *Polit. Theory* **25**, 347–376 (1997).
37. Wu, S., Hofman, J. M., Mason, W. A. & Watts, D. J. Who says what to whom on Twitter. *Proc. 20th International Conference on World Wide Web (WWW '11)* 705–714 (ACM, 2011).
38. Eady, G., Nagler, J., Guess, A., Zalinsky, J. & Tucker, J. A. How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *SAGE Open* https://doi.org/10.1177/2158244019832705 (2019).
39. Guess, A. (Almost) everything in moderation: new evidence on Americans' online media diets. *Am. J. Polit. Sci.* **65**, 1007–1022 (2020).
40. King, G. & Persily, N. A new model for industry–academic partnerships. *PS Polit. Sci. Polit.* **53**, 703–709 (2019).
41. Lazer, D. M. J. et al. Computational social science: obstacles and opportunities. *Science* **369**, 1060–1062 (2020).
42. Mynatt, E. et al. *Harnessing the Computational and Social Sciences to Solve Critical Social Problems* Tech. Rep. (National Science Foundation, 2020).
43. Hosseinmardi, H. et al. Examining the consumption of radical content on YouTube. *Proc. Natl Acad. Sci.* **118**, e2101967118 (2021).
44. Allcott, H., Braghieri, L., Eichmeyer, S. & Gentzkow, M. The welfare effects of social media. *Am. Econ. Rev.* **110**, 629–76 (2020).
45. Ansolabehere, S., Rodden, J. & Snyder, J. M. The strength of issues: using multiple measures to gauge preference stability, ideological constraint, and issue voting. *Am. Polit. Sci. Rev.* **102**, 215–232 (2008).
46. Dias, N. & Lelkes, Y. The nature of affective polarization: disentangling policy disagreement from partisan identity. *Am. J. Polit. Sci.* **66**, 775–790 (2021).
47. Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M. & Ryan, J. B. Affective polarization, local contexts and public opinion in America. *Nat. Hum. Behav.* **5**, 28–38 (2021).
48. Cohen, G. L. Party over policy: the dominating impact of group influence on political beliefs. *J. Pers. Soc. Psychol.* **85**, 808–822 (2003).
49. Arceneaux, K & Wielen, R. J. V. *Taming Intuition: How Reflection Minimizes Partisan Reasoning and Promotes Democratic Accountability* (Cambridge Univ. Press, 2017).
50. Angrist, J. D., Imbens, G. W. & Rubin, D. B. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* **91**, 444–455 (1996).
51. Yeomans, M., Kantor, A. & Tingley, D. The politeness package: detecting politeness in natural language. *R J.* **10**, 489–502 (2018).
52. Yeomans, M., Minson, J., Collins, H., Chen, F. & Gino, F. Conversational receptiveness: improving engagement with opposing views. *Organ. Behav. Hum. Decis. Process.* **160**, 131–148 (2020).
53. Hartman, R. et al. Interventions to reduce partisan animosity. *Nat. Hum. Behav.* **6**, 1194–1205 (2022).
54. Levendusky, M. S. & Malhotra, N. Does media coverage of partisan polarization affect political attitudes? *Polit. Commun.* **33**, 283–301 (2016).
55. Ahler, D. J. & Sood, G. The parties in our heads: misperceptions about party composition and their consequences. *J. Polit.* **80**, 964–981 (2018).
56. Moore-Berg, S. L., Ankori-Karlinsky, L.-O., Hameiri, B. & Bruneau, E. Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proc. Natl Acad. Sci. USA* **117**, 14864–14872 (2020).
57. Paluck, E. L., Green, S. A. & Green, D. P. The contact hypothesis re-evaluated. *Behav. Public Policy* **3**, 129–158 (2019).
58. Enders, A. M. & Armaly, M. T. The differential effects of actual and perceived polarization. *Polit. Behav.* **41**, 815–839 (2019).
59. Ruggeri, K. et al. The general fault in our fault lines. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-021-01092-x (2021).
60. Rossiter, E. The consequences of interparty conversation on outparty affect and stereotypes. In *2020 Meeting of the Society for Political Methodology* (2020).
61. Wojcieszak, M. & Warner, B. R. Can interparty contact reduce affective polarization? A systematic test of different forms of intergroup contact. *Polit. Commun.* https://doi.org/10.1080/10584609.2020.1760406 (2020).
62. Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M. & Ryan, J. B. (Mis-)estimating affective polarization. *J. Polit.* https://doi.org/10.1086/715603 (2021).

63. White, I. K. & Laird, C. N. *Beyond Adversary Democracy* (Princeton Univ. Press, 2020).
64. Grossmann, M. & Hopkins, D. A. *Asymmetric Politics: Ideological Republicans and Group Interest Democrats* 1st edn (Oxford Univ. Press, 2016).
65. Jahani, E. et al. An online experiment during the 2020 US–Iran crisis shows that exposure to common enemies can increase political polarization. *Sci. Rep.* **12**, 19304 (2022).
66. Friedkin, N. & Johnsen, E. Social influence networks and opinion change. *Adv. Group Process.* **16**, 1–29 (1999).

Acknowledgements

We thank C. Goode for assistance with software development. This research was funded by the Provost's Office at Duke University (C.A.B., D.S.H. and A.V.), a Facebook Foundational Research Award (C.A.B. and A.V.), Templeton Foundation Award No. 62656 (C.A.B., D.S.H. and A.V.) and NSF CAREER Award No. DMS-2046880 (A.V.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.V., D.S.H., C.A.B., A.C., G.T., F.M. and B.G. designed the research. A.C. and C.A.B. created the mobile communication platform. B.G., D.S.H. and C.A.B. designed the survey. G.T., A.V. and C.A.B. analysed the data. D.S.H., C.A.B., A.V., A.C., B.G. and G.T. wrote the manuscript.

Competing interests

C.A.B. served as an academic consultant for Twitter's Incentives Team in early 2022, which explored new ways to increase positive behaviour on its platform. In this capacity, he has been paid US$2,675. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01655-0.

Correspondence and requests for materials should be addressed to Alexander Volfovsky.

Peer review information *Nature Human Behaviour* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Alexander Volfovsky

Last updated by author(s): Mar 29, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were collected via the DiscussIt application and stored via Google Firebase.
Data analysis	All data analysis was performed in R, version 4.2.1 (using packages sandwich, ivpack, and lmtest), and Python version 3.9.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Upon publication, anonymized replication data will be made publicly available by the authors at this link: https://doi.org/10.7910/DVN/LTVEHJ

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We do not perform sex or gender based analysis.
Population characteristics	See below study description and Supplemental Appendix Section 2.3 and 2.4.
Recruitment	See Supplemental Appendix Section 2.2 for a detailed description. We hired the survey research firm YouGov to recruit a pool of potential respondents who were U.S. citizens at least 18 years of age, self-identified as either Republican or Democrat (including Independents who said they `leaned' toward one party), used an iOS or Android smartphone or tablet, and self-reported a willingness to install an app on their phone or tablet. We note that the original YouGov sample is self-selected and so may be more computer-literate than the general population as they are participating in online research. Supplement Table 1 and associated discussion provides details on how our recruited sample compares to national averages in terms of demographics. We also discuss how the recruited sample differs from the larger YouGov sample in terms of willingness to participate in social-media research. Our findings largely mirror those of previous research that younger, more politically knowledgeable and educated individuals are more willing to participate in such a study. We provide detailed analysis of this in Supplement Table 2 and subsequently adjust for covariates in our analyses.
Ethics oversight	The Duke University IRB reviewed and approved this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The randomized experiment provided treatment respondents access to a mobile app that paired Democrats with Republicans to have conversations about gun control or immigration. Respondents answered detailed questions about policy preference and attitudes in two ostensibly unrelated pre- and post-conversation surveys administered by YouGov. Quantitative measures of political polarization were compared before and after the conversation.
Research sample	We hired the survey research firm YouGov to recruit a pool of potential respondents who were U.S. citizens at least 18 years of age, self-identified as either Republican or Democrat (including Independents who said they `leaned' toward one party), used an iOS or Android smartphone or tablet, and self-reported a willingness to install an app on their phone or tablet. The non-probability sample has been selected to match the population of interest but its representativeness along other, unobserved, characteristics is unknown. We want to study how conversation with someone from a different political party would affect views towards that party and that party's policy preferences, so a sample of partisan Americans was required. The sample is slightly older, more female, and more educated than the US population as a whole. See Supplemental Appendix Section 2.3 for a detailed discussion.
Sampling strategy	YouGov used its quota-based online national sample to recruit participants, stratified by YouGov such that we had equal numbers of Democrat and Republican participants.
Data collection	The application DiscussIt logged all messages and survey responses across all experimental conditions. YouGov administered the pre- and post-surveys on-line. No one except the participants and researchers could see the messages and survey responses. Initial power calculations indicated that we needed 250 observations in each treatment arm to detect effect sizes of 0.25 standard deviations with 80% power. Data collection (both survey and within-app data) were blinded from the researcher.
Timing	Recruitment began on January 24, 2020 and data collection ended on February 27, 2020.
Data exclusions	Data were excluded if participants did not express willingness to download an app. Details are provided in Supplementary Appendix Section 2.2.
Non-participation	All individuals who completed the post-survey (where outcomes were measured) are used in the analysis and two-stage least-squares was used to estimate complier average causal effects. We have a completion rate of 22% of the 7074 individuals who were recruited by YouGov and consented to participate in the study. Details are provided in Supplementary Appendix Section 2.2.

Randomization	Participants were randomized into experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging