DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

Classification: Social Sciences (Major) and Psychological and Cognitive Sciences (Minor)

Title: Interventions Reducing Affective Polarization Do Not Improve Anti-Democratic Attitudes

Authors: Jan G. Voelkel¹, James Chu¹, Michael N. Stagnaro¹², Joseph S. Mernyk¹, Chrystal Redekopp¹, Sophia L. Pink¹, James N. Druckman³, David G. Rand², and Robb Willer¹

¹: Department of Sociology, Stanford University, Stanford, USA
²: Sloan School of Management, Massachusetts Institute of Technology, Cambridge, USA
³: Department of Political Science, Northwestern University, Evanston, USA

Corresponding authors: Robb Willer and Jan G. Voelkel, Department of Sociology, Stanford University, Stanford, CA 94305, USA. Contact: willer@stanford.edu and jvoelkel@stanford.edu.

Funding Statement: The authors received funding for this project from the Civic Health Project, the Stanford Center on Philanthropy and Civil Society, and the Institute for Policy Research, Northwestern University.

Author Contributions: JGV, JC, MNS, JSM, CR, SLP, JND, DGR, and RW designed the studies. JGV, JC, MNS, JSM, CR, SLP, and RW collected the data. JGV analyzed the data. JGV and DGR wrote the manuscript. JC, MNS, JND, and RW provided comments on the manuscript.

Competing Interests: The authors declare no competing interests.
Abstract

There is widespread concern that rising affective polarization – dislike for members of the opposing party – is exacerbating a range of anti-democratic attitudes, such as support for undemocratic practices, undemocratic candidates, and partisan violence. Accordingly, scholars and practitioners alike have invested great effort in developing depolarization interventions, and several promising interventions have been identified that successfully reduce affective polarization. Critically, however, it remains unclear whether these interventions reduce consequential anti-democratic attitudes, versus merely changing sentiments toward outpartisans.

In this paper, we address this question with high-power experimental tests of two previously established depolarization interventions, one a correction of misperceptions of the views of outpartisans, the other an invitation to think about interpartisan friendships (total \(n = 4,512\)). We find that the depolarization interventions reliably reduce affective polarization, but this reduction does not reliably translate into reduced support for undemocratic practices, undemocratic candidates, or partisan violence. Thus, efforts to strengthen pro-democratic attitudes should target these outcomes directly, rather than following the current practice of focusing on affective polarization as a proxy. More broadly, these findings call into question the previously assumed causal link of affective polarization on anti-democratic attitudes.
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

Significance Statement

The tendency to dislike supporters of the other party - affective polarization - is widely viewed as a threat to democratic societies. Accordingly, academics and civil society organizations have sought to develop interventions that reduce antipathy towards opposing partisans. Our research demonstrates a substantial limitation of such work: two interventions that reliably reduce affective polarization do not in fact reduce anti-democratic attitudes. Thus, future efforts should focus on developing new interventions that directly target anti-democratic attitudes, rather than using dislike of opposing partisans as a proxy. Furthermore, our findings suggest that affective polarization may not be as problematic for democratic societies as is widely assumed.
Affective polarization – the tendency of partisans to view opposing partisans negatively and copartisans positively (1) – has been a major focus of research in recent years (2, 3). In particular, research shows that contemporary U.S. politics is characterized by growing affective polarization (4, 5). Notably, not only academics but also a large majority of Americans believe the country is extremely divided (6) and view this division as a serious problem (7).

There is great concern about rising affective polarization because its presumed negative consequences may be uniquely harmful or destabilizing for democratic societies - for example by stimulating support for undemocratic candidates and practices, or by fomenting political violence (e.g., 2, 8-13). In light of the presumed dire consequences of affective polarization, academics and practitioners have invested a great deal of energy in developing interventions that reduce affective polarization, typically using outcomes based on sentiment towards opposing partisans (e.g., 12, 14-25). This body of work has uncovered numerous effective approaches for reducing affective polarization, tools that offer hope for maintaining – or restoring – democratic norms and practices.

Unfortunately, this hope may be premature. This is because nearly all prior work has focused on treating affective polarization itself, and assumed that reductions of affective polarization would in turn improve downstream outcomes that pose consequential threats to democracy or what Finkel et al. refer to as “dark consequences” (2, p. 535). Although this assumption may seem reasonable, there is little evidence evaluating its implications for the benefits of depolarization. Worryingly, a recent working paper found that manipulating affective
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

polarization using a trust game had no significant causal effect on accountability or attitudes about democratic norms (e.g., overriding elections, violence), despite significant observational associations (26). Thus, it remains unclear whether commonly used depolarizing interventions would have the hoped-for consequence of effectively reducing anti-democratic attitudes.

Here, we shed light on this question by testing the impact of two previously-proposed interventions on a variety of anti-democratic attitudes. In doing so, we advance the literature on affective polarization in three ways. First, we assess the replicability of prior findings by testing the interventions’ effect on standard sentiment measures of affective polarization. Second, we extend prior findings by testing the interventions’ effects on incentivized behavioral measures of affective polarization to address concerns that standard sentiment measures of affective polarization are inconsequential partisan signaling (e.g., 1, 27, 28).

Third, and most importantly, we test whether the effects of these interventions extend beyond affective polarization to impact the more societally-consequential outcomes of support for undemocratic practices, undemocratic candidates, and partisan violence. If results show that existing depolarization interventions reduce these more consequential outcomes, this would suggest that depolarization researchers and practitioners should extend their current work toward identifying the mechanisms underlying the effects on affective polarization, estimating the durability of these effects, and comparing the effectiveness of future ideas to the effectiveness of current interventions. However, if we find that existing depolarization interventions do not reduce anti-democratic attitudes, this would suggest that, to speak to these issues, depolarization research would need a fundamental realignment, including the development of new interventions and direct measurement of anti-democratic outcomes.
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

Results

We examined whether existing depolarization interventions do not only reduce affective polarization but also anti-democratic attitudes in two large-scale experiments. Study 1 \((n = 2,451)\) was conducted on a nationally representative sample, and Study 2 \((n = 2,061)\) followed a pre-registered analysis script.

Correlational Results

We begin by assessing the correlations between affective polarization and anti-democratic attitudes. In both studies, we measured affective polarization with its most common measure, a feeling thermometer that indicates how cold participants felt towards outpartisans. In addition, we measured two behavioral indicators of dislike for outpartisans in Study 1. First, we measured how much money participants would give to an outpartisan in a dictator game (reverse-coded). Second, we measured how much money participants would spend to take money away from an outpartisan in a “joy of destruction” game (29). Finally, we measured several anti-democratic attitudes. We measured support for undemocratic practices by copartisans with five (Study 1) or three (Study 2) items, e.g., “[Democrats/Republicans] should redraw districts to maximize their potential to win more seats in federal elections, even if it may be technically illegal” (18). We measured support for undemocratic inparty candidates with six (Study 1) or four (Study 2) items, e.g., “How likely would you be to vote for the [Democratic/Republican] candidate if you learned that they support a proposal to reduce the number of polling stations in areas that support the [Republican/Democratic] party?” (e.g., 30). We measured support for partisan violence with four items, e.g., “How much do you feel it is justified for [Democrats/Republicans] to use violence in advancing their political goals these
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

days?" (31). Prior research considers such anti-democratic attitudes as direct consequences of affective polarization, leading to assumptions that interventions that reduce affective polarization should also reduce these anti-democratic attitudes (e.g., 2).

The correlational evidence does not align with the commonly-held assumption that affective polarization leads to anti-democratic attitudes. In Study 1, affective polarization was weakly to moderately correlated with withholding money in a dictator game ($r = .34, p < .001$) and weakly correlated with spending money in a joy of destruction game ($r = .07, p < .001$). Affective polarization was weakly correlated with support for undemocratic practices ($r = .05, p = .011$) and undemocratic politicians ($r = .19, p < .001$), and negatively correlated with support for partisan violence ($r = -.22, p < .001$). In Study 2, affective polarization was weakly correlated with support for undemocratic practices ($r = .16, p < .001$) and undemocratic politicians ($r = .26, p < .001$), and was uncorrelated with support for partisan violence ($r = .02, p = .463$). Here we do not adjust for multiple comparisons – doing so would make the associations between affective polarization and downstream measures even more tenuous.

Experimental Results

We now turn to our main focus, testing the causal effects of promising depolarization interventions on both (i) measures used in prior research and (ii) measures of potential downstream consequences for democracy. In Study 1, we tested two recently proposed interventions – one in which an outparty friendship was made salient (12), with the logic being that thinking of an outparty friend will generate more positive or less threatening general feelings about the other party. The second intervention involves correcting exaggerated misperceptions about the extent of outparty opposition to inparty collective political behaviors (16; see 14 for a
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

similar intervention). This misperception correction intervention makes clear the other party is not as much of a threat to one’s party agenda as if often believed; it has been replicated across 25 countries (19). Both interventions are compared to a control group. Afterwards, we measured the variables described above for the correlational analyses: affective polarization, withholding money from an outpartisan in a dictator game, spending money to take money away from an outpartisan in a joy of destruction game, support for undemocratic practices by the copartisans, support for undemocratic inparty candidates, and support for partisan violence.

The main findings are shown in Figure 1. First, both interventions reduced affective polarization as measured by cold feelings toward outpartisans. Participants in the friendship intervention condition \((M = 69.9, \ SD = 24.7)\) and in the misperception correction intervention condition \((M = 70.3, \ SD = 24.0)\) reported significantly lower levels of affective polarization than participants in the control condition \((M = 74.1, \ SD = 24.3)\) – thus replicating prior results (for the friendship intervention condition: \(b = -3.86, \ SE = 1.20, \ t(2326) = -3.20, \ p = .001, \ Cohen’s \ d = -0.16\); for the misperception correction intervention condition: \(b = -3.66, \ SE = 1.20, \ t(2326) = -3.04, \ p = .002, \ Cohen’s \ d = -0.15\)).

Second, both interventions also reduced behavioral indicators of affective polarization. Participants in the friendship intervention condition \((M = 32.9, \ SD = 14.2)\) and in the misperception correction intervention condition \((M = 32.3, \ SD = 13.6)\) withheld significantly less money from an outpartisan in a dictator game than participants in the control condition \((M = 35.0, \ SD = 14.2)\) (for the friendship intervention condition: \(b = -2.14, \ SE = 0.70, \ t(2326) = -3.07, \ p = .002, \ Cohen’s \ d = -0.15\); for the misperception correction intervention condition: \(b = -2.75, \ SE = 0.70, \ t(2326) = -3.94, \ p < .001, \ Cohen’s \ d = -0.20\)). Participants in the friendship
intervention condition ($M = 8.0, SD = 9.8$) also spent significantly less money on taking money away from an outpartisan in a joy of destruction game than participants in the control condition ($M = 9.1, SD = 10.1$) ($b = -1.12, SE = 0.48, t(2316) = -2.32, p = .020$, Cohen’s $d = -0.11$). The misperception correction intervention ($M = 8.3, SD = 9.4$) did not significantly reduce spending in the joy of destruction game ($b = -0.74, SE = 0.48, t(2316) = -1.53, p = .127$, Cohen’s $d = -0.08$).

Critically, however, neither intervention significantly reduced any of the downstream measures, even without correcting for multiple comparisons. Participants in the friendship intervention condition ($M = 37.2, SD = 25.8$) and in the misperception correction intervention condition ($M = 35.0, SD = 24.9$) did not report significantly less support for undemocratic practices by copartisans than participants in the control condition ($M = 37.3, SD = 25.3$) (for the friendship intervention condition: $b = -0.22, SE = 1.18, t(2319) = -0.19, p = .851$, Cohen’s $d = -0.01$; for the misperception correction intervention condition: $b = -1.71, SE = 1.18, t(2319) = -1.46, p = .145$, Cohen’s $d = -0.07$). Participants in the friendship intervention condition ($M = 55.9, SD = 21.1$) and in the misperception correction intervention condition ($M = 54.2, SD = 20.6$) did not report significantly less support for undemocratic inparty candidates than participants in the control condition ($M = 55.2, SD = 21.9$) (for the friendship intervention condition: $b = 0.47, SE = 0.99, t(2263) = 0.48, p = .631$, Cohen’s $d = 0.02$; for the misperception correction intervention condition: $b = -0.75, SE = 0.98, t(2263) = -0.76, p = .447$, Cohen’s $d = -0.04$). Participants in the friendship intervention condition ($M = 15.1, SD = 25.7$) and in the misperception correction intervention condition ($M = 13.2, SD = 22.8$) did not report significantly less support for partisan violence than participants in the control condition ($M =
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

15.6, \( SD = 25.1 \) (for the friendship intervention condition: \( b = -0.84, SE = 1.15, t(2281) = -0.73, p = .468 \), Cohen’s \( d = -0.03 \); for the misperception correction intervention condition: \( b = -1.52, SE = 1.15, t(2281) = -1.32, p = .185 \), Cohen’s \( d = -0.06 \)). Taken together, the results from Study 1 show that interventions can reduce both attitudinal and behavioral indicators of affective polarization without reducing anti-democratic attitudes. This calls into question the commonly-held assumption that anti-democratic attitudes are downstream consequences of affective polarization.

In Study 2 we sought to address an important limitation of Study 1: the lack of full randomization of the order of the dependent variables. In Study 1, support for partisan violence and support for undemocratic candidates were always presented after the feeling thermometer, the two behavioral measures of affective polarization and the support for undemocratic practices measure. Thus, the lack of effects on support for partisan violence and support for undemocratic candidates could have been because the effect of the treatment decreased over time (due to participant fatigue or something else). In Study 2, we fully randomized the order of four dependent variables: the feeling thermometer, support for undemocratic practices, support for undemocratic candidates, and support for partisan violence. We focused on the comparison of the control condition and the misperception correction intervention, as this intervention showed larger (yet statistically non-significant) effects on anti-democratic attitudes than the friendship intervention in Study 1. We did not include the behavioral measures of affective polarization in Study 2, given their secondary interest.

The results of Study 2 were similar to the results of Study 1 (see Figure 2). Once again, participants in the misperception correction intervention condition (\( M = 76.7, SD = 21.0 \))
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

reported significantly lower levels of affective polarization than participants in the control condition \( (M = 80.1, SD = 19.7) \) \( (b = -2.90, SE = 0.88, t(2007) = -3.29, p = .001, \text{Cohen’s } d = -0.14) \). However, even without correcting for multiple comparisons, the misperception correction intervention did not significantly reduce any of the negative downstream outcomes. Participants in the misperception correction intervention condition \( (M = 16.2, SD = 21.2) \) did not report significantly less support for undemocratic practices by copartisans than participants in the control condition \( (M = 17.8, SD = 21.6) \) \( (b = -1.42, SE = 0.93, t(2004) = -1.53, p = .127, \text{Cohen’s } d = -0.07) \). Participants in the misperception correction intervention condition \( (M = 51.0, SD = 19.4) \) did not report significantly less support for undemocratic inparty candidates than participants in the control condition \( (M = 50.3, SD = 20.6) \) \( (b = 0.90, SE = 0.86, t(1998) = 1.05, p = .295, \text{Cohen’s } d = 0.04) \). Participants in the misperception correction intervention condition \( (M = 6.7, SD = 16.0) \) did not report significantly less support for partisan violence than participants in the control condition \( (M = 7.1, SD = 15.6) \) \( (b = -0.37, SE = 0.69, t(2002) = -0.54, p = .590, \text{Cohen’s } d = -0.02) \).

To provide higher-power \( (n = 3,700) \) tests of the effect of the misperception correction intervention on anti-democratic attitudes, we also pooled the data from the misperception correction intervention condition and the control condition from Studies 1 and 2. However, the misperception correction intervention did not significantly reduce support for undemocratic inparty candidates (Cohen’s \( d = 0.00, p = .919 \)) or support for partisan violence (Cohen’s \( d = -0.05, p = .089 \)) relative to the control group. The effect of the misperception correction intervention on support for undemocratic practices by copartisans was statistically significant.
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

(Cohen’s $d = -0.07$, $p = .022$), but this difference is quite small in magnitude and was not robust to a Holm-Bonferroni adjustment for multiple hypothesis testing.

In both studies, the treatment effects were similar for Democratic and Republican participants. Separate test statistics for these two groups are provided in the Supporting Information. In Study 1, we found some evidence that the effects of the misperception correction intervention on affective polarization was stronger for Democrats than for Republicans, but this effect was not replicated in Study 2.

**Discussion**

Our findings call into question whether depolarization interventions developed to reduce affective polarization also reduce anti-democratic attitudes. Across two highly powered experiments, we successfully replicate previous findings of two depolarization interventions reliably reducing self-reported affective polarization. We also extend past work by showing that these effects extend to behavioral indicators of affective polarization, thereby demonstrating that depolarization interventions can impact behaviors with real, monetary stakes for outpartisans. Critically, however, the depolarization interventions did not reliably reduce support for undemocratic practices by copartisans, support for undemocratic inparty candidates, or support for partisan violence. Even the correlational associations between affective polarization and these anti-democratic attitudes were weak or non-existent. Thus, we conclude that many researchers (including ourselves; e.g., 2) may have substantially overestimated the societal benefits of depolarization interventions.

Our results suggest that research on depolarization interventions needs a fundamental realignment. Future research should focus on the development of depolarization interventions
that impact the outcomes of depolarization that are deemed most important. That is, researchers and practitioners who are interested in interventions targeting anti-democratic attitudes such as support for undemocratic practices, undemocratic politicians, and support for partisan violence should not focus on treating affective polarization and begin developing more direct interventions – trends that run counter to most current work.

From a broader theoretical perspective, our results raise serious questions about whether a causal link from affective polarization to antidemocratic attitudes actually exists. This has two theoretical implications. First, the causal forces that have been identified as leading to affective polarization (e.g., social and ideological sorting, partisan media, elite ideological polarization) may or may not be connected to the democratic backsliding that scholars in the US have identified (e.g., 30, 32-34). Second, it is unclear whether affective polarization itself – even at high levels – poses a democratic threat. The psychological basis for anti-democratic attitudes involves factors distinct from affective polarization and the two should not be confused. Identifying factors that lead to anti-democratic attitudes should be a priority for future research (e.g., 32).

Methods

Ethics Statement and Reproducibility

All studies were approved by the Institutional Review Board at Stanford University. All participants provided informed consent. Materials, anonymized data (including descriptions of how the original files were anonymized), and analysis code for both studies as well as the preregistration for Study 2 are available via https://osf.io/n5u9d/?view_only=bd46d6d6d32e4a43ac67130639788280.
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

Samples

In Study 1, we collected data from 2,451 participants who were recruited from an Internet panel provided by Bovitz Inc between October 28, 2020 to November 3, 2020. Bovitz maintains an online panel of approximately one million respondents recruited through random digit dialing and empanelment of Americans with Internet access. Samples are drawn such that the demographics of the sample match those of the U.S. population (as measured by the American Community Survey). In Study 2, we collected data from 2,061 participants who were recruited from a large panel of previously recruited Amazon Mechanical Turk workers between March 23, 2021 and April 5, 2021. Participants were excluded if they had duplicate IDs (keeping only the first case), did not consent to participate, were underage or did not provide their age, failed multiple attention checks (completed pre-treatment-assignment), identified as neither Democrat nor Republican, or left the study before they were randomly assigned to a condition. We used pairwise deletion for all dependent variables. The sample was recruited with soft quotas for participants’ self-identified partisanship (including learners): 50% Democrats and 50% Republicans. Both studies had large sample sizes. According to sensitivity power analyses conducted with G*Power (35), both studies had at least 95% power to detect even small effect sizes (Cohen’s $d > 0.17$ in Study 1 and Cohen’s $d > 0.15$ in Study 2). More details on sample characteristics, exclusion rules, and demographics are provided in the Supporting Information.

Interventions

Participants were randomly assigned to one of three (Study 1) or two (Study 2) conditions. In the friendship intervention condition (12) which was only included in Study 1, participants received the following instructions: “Although you are [a Democrat/an Independent
who is closer to the Democratic Party/an Independent who is closer to the Republican Party/a Republican], you likely know people who are [Republicans/Democrats]. Think about one such [Republican/Democrat] that you like and respect a great deal. This person could be a friend, relative, neighbor, co-worker, or just someone that you know. Please explain why you feel this way about this person.” In the misperception correction intervention condition (16), participants were presented with a scenario, randomly chosen from five scenarios, e.g., “A state [Democratic/Republican] party in control of the state legislature has drafted a proposal to streamline the appointment of judges where judges would be nominated and voted on in groups, not individually. This would reduce the workload of state legislators and make the process more efficient, however it may make it more difficult for the party in the minority, the [Republicans/Democrats], to object to the appointment of individual judges” (all scenarios are available at https://osf.io/n5u9d/?view_only=bd46d6d6d32e4a43ac67130639788280). Then participants rated how much they believed an outpartisan would (a) dislike and (b) oppose this action, and (c) find this action politically unacceptable. After partisans provided their own beliefs, their potential misperceptions were corrected by presenting the real responses from outpartisans and the real responses from inpartisans who had a read a similar scenario where outpartisans were taking the action. The real responses were based on a nationally representative survey (16). In the control condition, participants moved immediately towards the section with the measures of the dependent variable.

**Measures**

In both studies, we measured *affective polarization* using a feeling thermometer: “We would like to get your feelings toward both Democrats and Republicans. We would like you to
rate them using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward them. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward them and that you don't care too much for them. You would rate them at the 50 degree mark if you don't feel particularly warm or cold toward them.” We used the reversed-coded feeling thermometer toward outpartisans.

*Withholding money in a dictator game* was measured with the following item: “You have been anonymously and randomly matched with another participant who identifies as a [Republican/Democrat]. You have just been given 50 cents. You will now decide how to split these 50 cents between yourself and the [Republican/Democratic] participant. You can give any amount between 0 cents and 50 cents to the other participant. The other participant cannot affect the outcome you choose.”

*Spending money in a joy-of-destruction game* was measured with the following item (based on 29): “You have been anonymously and randomly matched with another participant who identifies as a [Republican/Democrat]. Both you and the other participant have just each been given 50 cents. You will now decide whether to leave the [Republican/Democratic] participant’s payment unchanged or take away part or all of their 50 cents. For every 1 cent you pay, you remove 2 cents from the [Republican/Democratic] participant. You can pay any amount between 0 cents and 25 cents. The other participant cannot affect the outcome you choose.”

*Support for undemocratic practices by copartisans* was measured on a 101-point scale ranging from “strongly disagree” to “strongly agree” with the following items (18): (i) I think the [Democrats/Republicans] should do everything they can to hurt the [Republican/Democratic] party, even if it is at the short-term expense of the country, (ii) It’s OK to sacrifice US economic
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

prosperity in the short term in order to hurt [Republicans'/Democrats’] chances in future elections, (iii) [Democrats/Republicans] should redraw districts to maximize their potential to win more seats in federal elections, even if it may be technically illegal, (iv) If [Democrats/Republicans] gain control of all branches of government, they should use the Federal Communications Commission to heavily restrict or shut down [Fox News/MSNBC] to stop the spread of fake news, and (v) I think the [Democrats/Republicans] should do everything in their power within the law to make it as difficult as possible for [Trump to run the government effectively/Democrats to take part in governing the country]. Items iv and v were not included in Study 2.

Support for undemocratic inparty candidates was measured on a 101-point scale ranging from “extremely likely to vote for the [Republican/Democratic] candidate” to “extremely likely to vote for the [Democratic/Republican] candidate” with the following items (based on 30): (i) How likely would you be to vote for the [Democratic/Republican] candidate if you learned that they said they would ignore unfavorable court rulings by [Republican/Democratic]-appointed judges?, (ii) How likely would you be to vote for the [Democratic/Republican] candidate if you learned that they support a proposal to reduce the number of polling stations in areas that support the [Republican/Democratic] party?, (iii) How likely would you be to vote for the [Democratic/Republican] candidate if you learned that they support a redistricting plan that gives [Democrats/Republicans] 10 extra seats despite a decline in the polls?, (iv) How likely would you be to vote for the [Democratic/Republican] candidate if you learned that they said that [Democrats/Republicans] should not accept election results if they do not win?, (v) How likely would you be to vote for the [Democratic/Republican] candidate if you learned that they said
they would prosecute journalists who accuse them of misconduct if the journalists won’t reveal their sources?, and (vi) How likely would you be to vote for the [Democratic/Republican] candidate if you learned that they said they would ban far-[right/left] group rallies on the state capital grounds?. Items v and vi were not included in Study 2.

Support for partisan violence was measured with the following four items (31): (i) When, if ever, is it OK for $\text{Inparty\_Person}$s to send threatening and intimidating messages to $\text{Outparty\_Party}$ party leaders?, (ii) When, if ever, is it OK for an ordinary $\text{Inparty\_Person}$ in the public to harass an ordinary $\text{Outparty\_Person}$ on the Internet, in a way that makes the target feel frightened?, (iii) How much do you feel it is justified for $\text{Inparty\_Person}$s to use violence in advancing their political goals these days?, and (iv) How much do you feel it is justified for $\text{Inparty\_Person}$s to use violence if the $\text{Outparty\_Party}$ party wins more races in the next election?. Items i and ii used a 101-point scale ranging from “never” to “always” and items iii and iv used a 101-point scale ranging from “not at all justified” to “extremely justified”.

In both studies, we included additional dependent variables that are not relevant for answering the research questions of this paper. The questionnaires for both studies are available via https://osf.io/n5u9d/?view_only=bd46d6d6d32e4a43ac67130639788280.
DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES

Figure 1: Effects of the Friendship Intervention and the Misperception Correction Intervention on Affective Polarization and Anti-Democratic Attitudes, Study 1
Figure 2: Effects of the Misperception Correction Intervention on Affective Polarization and Anti-Democratic Attitudes, Study 2
References


DEPOLARIZATION INTERVENTIONS AND ANTI-DEMOCRATIC ATTITUDES


   https://www.wm.edu/offices/global-research/_documents/pips/selene-swanson-whitepaper


   https://osf.io/9btsq/download


