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

Author list: Jan G. Voelkel[1*], James Chu[2], Michael N. Stagnaro[3], Joseph S. Mernyk[1], Chrystal Redekopp[1], Sophia L. Pink[4], James N. Druckman[5], David G. Rand[3], and Robb Willer[1*]

Affiliations:

[1]: Department of Sociology, Stanford University, Stanford, USA

2: Department of Sociology, Columbia University, New York, USA

3: Sloan School of Management, Massachusetts Institute of Technology, Cambridge, USA

4: Wharton School, University of Pennsylvania, Philadelphia, USA

5: 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.

The paper waspublished at Nature Human Behaviour. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please cite as

Voelkel, J. G., Chu, J., Stagnaro, M. N., Mernyk, J. S., Redekopp, C., Pink, S. L., Druckman, J. N., Rand, D. G., & Willer, R. (2022). Interventions reducing affective polarization do not necessarily improve anti-democratic attitudes. *Nature Human Behaviour*.

https://doi.org/10.1038/s41562-022-01466-9

Abstract

There is widespread concern that rising affective polarization – particularly dislike for outpartisans – exacerbates Americans' anti-democratic attitudes. Accordingly, scholars and practitioners alike have invested great effort in developing depolarization interventions that reduce affective polarization. Critically, however, it remains unclear whether these interventions reduce anti-democratic attitudes, or only change sentiments toward outpartisans. In this paper, we address this question with experimental tests (total n=8,385) of three previously established depolarization interventions: correcting misperceptions of outpartisans, priming interpartisan friendships, and observing warm cross-partisan interactions between political leaders. While these depolarization interventions reliably reduced affective polarization, we do not find compelling evidence that these interventions reduced support for undemocratic candidates, support for partisan violence, or prioritizing partisan ends over democratic means. Thus, future efforts to strengthen pro-democratic attitudes may do better if they target these outcomes directly. More broadly, these findings call into question the previously assumed causal effect of affective polarization on anti-democratic attitudes.

Interventions Reducing Affective Polarization Do Not Necessarily Improve Anti-Democratic Attitudes

Affective polarization – the tendency of partisans to view opposing partisans negatively and co-partisans 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 in part 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 substantial energy in developing depolarization 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 much prior work has focused on treating affective polarization itself, and assumed that these interventions would in turn improve downstream outcomes that pose consequential threats to democracy (2). Although this assumption may seem reasonable, there is little evidence evaluating its implications for the benefits of depolarization interventions. Here, we shed light on the question of whether previously established depolarization interventions have the hoped-for consequence of effectively reducing anti-democratic attitudes.

Researchers who study depolarization interventions frequently propose that reducing affective polarization will indeed have positive effects on democratic outcomes. In line with this, ten of the twelve papers on depolarization interventions we identified in the literature discuss anticipated effects on democratic outcomes (see quotes in Supplementary Table 1). Recent review papers (2, 26) also discuss improving democratic outcomes as a goal of depolarization interventions.

The reason depolarization interventions are expected to reduce anti-democratic attitudes is that many researchers assume that affective polarization causes greater anti-democratic attitudes. The first link in this causal chain - the effect of depolarization interventions on affective polarization - is well supported empirically in the published literature. The second link - a causal effect of affective polarization on anti-democratic attitudes - has been suggested by many researchers (e.g., 2-3, 27-31), and there is little evidence that the relationship between affective polarization and anti-democratic attitudes is contested. In fact, with the exception of (32), we did not find any work suggesting a lack of a causal effect of affective polarization on anti-democratic attitudes.

There are several potential mechanisms through which reducing affective polarization could reduce anti-democratic attitudes. First, reducing affective polarization could reduce the perceived threat of the negative consequences when the outparty wins an election that, in turn, could reduce support for undemocratic in-party candidates. Second, reducing affective polarization could reduce identification with the inparty that, in turn, could decrease the desire to break democratic norms to win at all costs. Third, reducing affective polarization could increase empathy for outpartisans, making it more difficult to justify violence against them.

However, while there are many good reasons to expect that depolarization interventions reduce anti-democratic attitudes, to date there is little empirical evidence in favor of this hypothesis. One paper, using cross-sectional data, found that affective polarization is (significantly) negatively correlated with support for democratic norms (28). Conversely, a recent paper found that manipulating affective polarization had no significant causal effect on accountability, attitudes about democratic norms, or support for partisan violence (32). In short, it remains unclear whether commonly used depolarizing interventions would have the hoped-for consequence of reducing anti-democratic attitudes.

Here, we shed light on this question by testing the impact of three previously validated depolarization 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 robustness of prior findings by testing the interventions' effect on standard sentiment measure 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, with no impact on interpersonal behaviors (e.g., 1, 33-34).

Third, and most importantly, we test whether the effects of these interventions extend beyond affective polarization to impact three measures of anti-democratic attitudes: support for undemocratic candidates, support for partisan violence, and prioritizing partisan ends over democratic means. The outcomes we study were based on prior research on anti-democratic attitudes (18, 35-36) and chosen because of their relevance for contemporary US politics. 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 maximizing the effectiveness of current interventions. However, if we find that existing depolarization interventions do not reduce anti-democratic attitudes, this would suggest that reducing anti-democratic attitudes requires the development of new interventions and direct measurement of anti-democratic attitudes.

Results

We examined whether existing depolarization interventions not only reduce affective polarization but also anti-democratic attitudes in three large-scale experiments. Studies 1 (n = 2,341) and 3 (n = 4,023) were conducted on non-probability samples that were similar to the national population on several key demographics. Study 2 (n = 2,021) was conducted on a convenience sample. Studies 2 and 3 followed pre-registered analysis scripts.

Correlational Results

We begin by assessing the correlations between affective polarization and anti-democratic attitudes. In our three studies, we measured affective polarization with the canonical feeling thermometer item 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). Dictator games were used as a behavioral indicator in a seminal paper on affective polarization (1). Second, we measured how much money participants would spend to take money away from an outpartisan in a “joy of destruction” game (37). This behavioral indicator captures how much people are willing to personally sacrifice to reduce the earnings of outpartisans. 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).

Finally, we measured several anti-democratic attitudes. We measured support for undemocratic candidates - a willingness to sacrifice democratic principles in electoral contexts for inparty victories - with items such as “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?” (35). This approach follows work showing partisans are often willing to violate democratic norms (e.g., regarding electoral fairness, checks and balances, civil liberties) to win elections (35). We also measured support for partisan violence, another key facet of undemocratic attitudes, with items such as “How much do you feel it is justified for [Democrats/Republicans] to use violence in advancing their political goals these days?” (36). Finally, we measured prioritizing partisan ends over democratic means - a willingness to help the inparty at the expense of the country and/or in contravention of democratic norms - with items such as “[Democrats/Republicans] should redraw districts to maximize their potential to win more seats in federal elections, even if it may be technically illegal” (18).

Prior research considers such anti-democratic attitudes to be direct consequences of affective polarization, leading to assumptions that interventions that reduce affective polarization should also reduce these anti-democratic attitudes (e.g., 2). Accordingly, we expected that affective polarization would be correlated with anti-democratic attitudes. However, across our three studies, affective polarization was not reliably correlated with support for undemocratic politicians, support for partisan violence, or prioritizing partisan ends. Affective polarization was positively correlated with support for undemocratic politicians in Studies 1 (r = .19, p < .001,

$p_{\text{Holm-Bonferroni-Adjustment (HBA)}} < .001$) and 2 ($r = .26$, $p < .001$, $p_{HBA} < .001$) but negatively correlated in Study 3 ($r = -.05$, $p = .001$, $p_{HBA} = .001$). Affective polarization was positively correlated with prioritizing partisan ends over democratic means in Studies 1 ($r = .05$, $p = .011$, $p_{HBA} = .011$) and 2 ($r = .16$, $p < .001$, $p_{HBA} < .001$) but negatively correlated in Study 3 ($r = -.23$, $p < .001$, $p_{HBA} < .001$). Affective polarization was negatively correlated with support for partisan violence in Studies 1 ($r = -.22$, $p < .001$, $p_{HBA} < .001$) and 3 ($r = -.36$, $p < .001$, $p_{HBA} < .001$) and uncorrelated in Study 2 ($r = .02$, $p = .463$, $p_{HBA} = .463$). However, because the observed correlations could be suppressed by unobserved variables, we conducted experiments to estimate the causal effect of depolarization interventions on anti-democratic attitudes.

Experimental Results

We now turn to our main focus, testing the causal effects of promising depolarization interventions on both (i) measures of affective polarization and (ii) measures of anti-democratic attitudes. We chose a varied set of interventions that we perceived to be particularly promising ways to reduce affective polarization.

In Study 1, we tested two recently proposed interventions. The first intervention made an outparty friendship salient (12), following the logic that thinking of a friend who supports the opposing party will generate more positive and/or less threatening feelings about the other party. The second intervention corrected exaggerated misperceptions about the extent of outparty opposition to inparty attempts to pass a policy (16, 19; see 14 for a similar intervention). This misperception correction intervention makes clear the other party is less of a threat to the agenda of the participant's favored party as if often believed. Both interventions were compared to a control group. Afterward, we measured the variables described above in 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 candidates, support for partisan violence, and prioritizing partisan ends over democratic means. 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) (for the friendship intervention condition: b = -3.86, SE = 1.20, t(2326) = -3.20, p = .001, Cohen's d = -0.16, 95% CI for b = [-6.22, -1.49]; for the misperception correction intervention condition: b = -3.66, SE = 1.20, t(2326) = -3.04, p = .002, Cohen's d = -0.15, 95% CI for b = [-6.02, -1.30]).

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: b = -2.14, SE = 0.70, t(2326) = -3.07, p = .002, p_{HBA} = .004, Cohen's d = -0.15, 95% CI for b = [-3.51, -0.77]; for the misperception correction intervention: b = -2.75, SE = 0.70, t(2326) = -3.94, p < .001, p_{HBA} < .001, Cohen's d = -0.20, 95% CI for b = [-4.12, -1.38]). Participants in the friendship intervention condition (M = 8.0, SD = 9.8) also spent significantly less money to take 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, p_{HBA} = .020, Cohen's d = -0.11, 95% CI for b = [-2.07, -0.17]). 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, p_{HBA} = .127, Cohen's d = -0.08, 95% CI for b = [-1.68, 0.21]).

Critically, however, neither intervention significantly reduced any of the downstream measures. 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 candidates than participants in the control condition (M = 55.2, SD = 21.9) (for the friendship intervention: b = 0.47, SE = 0.99, t(2263) = 0.48, p = .631, p_{HBA} = 1, Cohen's d = 0.02, 95% CI for b = [-1.46, 2.41]; for the misperception correction intervention: b = -0.75, SE = 0.98, t(2263) = -0.76, p = .447, p_{HBA} = .447, Cohen's d = -0.04, 95% CI for b = [-2.68, 1.18]). 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 = 15.6, SD = 25.1) (for the friendship intervention: b = -0.84, SE = 1.15, t(2281) = -0.73, p = .468, p_{HBA} = 1, Cohen's d = -0.03, 95% CI for b = [-3.09, 1.42]; for the misperception correction intervention: b = -1.52, SE = 1.15, t(2281) = -1.32, p = .185, p_{HBA} = .436, Cohen's d = -0.06, 95% CI for b = [-3.78, 0.73]). 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 lower levels of prioritizing partisan ends over democratic means than participants in the control condition (M = 37.3, SD = 25.3) (for the friendship intervention: b = -0.22, SE = 1.18, t(2319) = -0.19, p = .851, p_{HBA} = 1, Cohen's d = -0.01, 95% CI for b = [-2.53, 2.09]; for the misperception correction intervention: b = -1.71, SE = 1.18, t(2319) = -1.46, p = .145, p_{HBA} = .436, Cohen's d = -0.07, 95% CI for b = [-4.02, 0.59]).

Bayesian analyses provided further evidence that neither intervention affected any of the downstream measures. Whereas non-significant p-values in null hypothesis significance testing cannot be interpreted as evidence in favor of null effects, the Bayes Factor quantifies the relative predictive performance of the null hypothesis of no difference relative to the alternative hypothesis (38). Bayesian analyses require a choice of the prior of the relative plausibility of the two hypotheses before looking at the data. Here, we used the prior that the null hypothesis and alternative hypothesis are equally likely to be true. Such a relatively high prior for the alternative hypothesis is justified because we and other researchers believed that depolarization interventions would reduce anti-democratic attitudes. The Bayes factor describes to what extent the data warrant a change in the relative plausibility of the two hypotheses from this prior (38).

We found strong evidence for null effects of the friendship intervention on support for undemocratic candidates ($BF_{01} = 14.06$), support for partisan violence ($BF_{01} = 16.43$), and prioritizing partisan ends over democratic means ($BF_{01} = 17.66$). We found strong evidence for a null effect of the misperception correction intervention on support for undemocratic candidates ($BF_{01} = 11.26$). The data were also more consistent with a null effect of the misperception correction intervention on support partisan violence ($BF_{01} = 2.88$) and prioritizing partisan ends over democratic means ($BF_{01} = 3.53$), but this evidence was weak and moderate respectively. Taken together, the results from Study 1 show that depolarization interventions can reduce both attitudinal and behavioral indicators of affective polarization without necessarily 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 a 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 prioritizing partisan ends over democratic means. 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 candidates, support for partisan violence, and prioritizing partisan ends over democratic means. 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) reported significantly lower levels of affective polarization than participants in the control condition (M = 80.2, SD = 19.5) (b = -2.90, SE = 0.88, t(2007) = -3.29, p = .001, Cohen's d = -0.14, 95% CI for b = [-4.63, -1.17]). However, the misperception correction intervention did not significantly reduce any of the negative downstream outcomes. Participants in the misperception correction intervention condition (M = 51.0, SD = 19.4) did not report significantly less support for undemocratic candidates than participants in the control condition (M = 50.2, SD = 20.5) (b = 0.90, SE = 0.86, t(1998) = 1.05, p = .295, p_{HBA} = .591, Cohen's d = 0.04, 95% CI for b = [-0.78, 2.57]). 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 = 15.4) (b = -0.37, SE = 0.69, t(2002) = -0.54, p = .590, p_{HBA} = .591, Cohen's d = -0.02, 95% CI for b = [-1.71, 0.98]). Participants in the misperception correction intervention condition (M = 16.1, SD = 21.1) did not report significantly lower levels of prioritizing partisan ends over democratic means than participants in the control condition (M = 17.7, SD = 21.4) (b = -1.42, SE = 0.93, t(2004) = -1.53, p = .127, p_{HBA} = .381, Cohen's d = -0.07, 95% CI for b = [-3.24, 0.40]).

Bayesian analyses provided further evidence that the misperception correction intervention did not affect the downstream measures. We found strong evidence for a null effect of the misperception correction intervention on support for undemocratic candidates (BF_{01} = 13.03) and on support for partisan violence (BF_{01} = 18.13). The data were also consistent with a null effect of the misperception correction intervention on prioritizing partisan ends over democratic means, but this evidence is only moderately strong: BF_{01} = 4.68). Taken together, the results from Study 2 replicate the finding that interventions reducing affective polarization do not necessarily reduce anti-democratic attitudes.

In Study 3, we tested an elite-focused intervention. While the friendship and misperception correction interventions tested in Studies 1 and 2 used content about outpartisan voters, Study 3 tested a warm elite relations intervention highlighting the friendship between a Republican politician (John McCain) and a Democratic politician (Joe Biden). This intervention is similar to the warm elite relations treatments used in prior research (15) but avoids deception.

The results of Study 3 again support the idea that interventions reduce affective polarization but do not reduce anti-democratic attitudes (see Figure 3). Participants in the warm elite relations intervention condition (M = 69.6, SD = 28.5) reported significantly lower levels of affective polarization than participants in the control condition (M = 72.0, SD = 27.6) (b = -2.04, SE = 0.83, t(4008) = -2.45, p = .014, Cohen's d = -0.07, 95% CI for b = [-3.67, -0.40]).

However, the warm elite relations intervention did not significantly reduce any of the measures of anti-democratic attitudes. On the contrary, participants in the warm elite relations intervention condition (M = 58.8, SD = 26.0) reported significantly more support for undemocratic candidates than participants in the control condition (M = 57.4, SD = 25.6) (b = 1.49, SE = 0.73, t(3954) = 2.04, p = .042, p_{HBA} = .083, Cohen's d = 0.06, 95% CI for b = [0.06, 2.93]; although this result was not robust to a Holm-Bonferroni adjustment for multiple hypothesis testing. Participants in the warm elite relations intervention condition (M = 24.4, SD = 32.1) also reported significantly more support for partisan violence than participants in the control condition (M = 22.0, SD = 31.0) (b = 2.20, SE = 0.88, t(3996) = 2.51, p = .012, p_{HBA} = .036, Cohen's d = 0.07, 95% CI for b = [0.48, 3.92]). Participants in the warm elite relations intervention condition (M = 34.4, SD = 31.8) did not report significantly lower levels of prioritizing partisan ends over democratic means than participants in the control condition (M = 33.4, SD = 31.5) (b = 0.76, SE = 0.90, t(3991) = 0.84, p = .401, p_{HBA} = .401, Cohen's d = 0.02, 95% CI for b = [-1.01, 2.53]).

Bayesian analyses provided further evidence that the warm elite relations intervention did not decrease the downstream measures. We found weak to moderate evidence for a null effect of the warm elite relations intervention on support for undemocratic candidates (BF_{01} = 6.18) and on support for partisan violence (BF_{01} = 1.52). Note that even if these effects were not null, the direction of these effects was positive, not negative. We found strong evidence for a null effect of the warm elite relations intervention on prioritizing partisan ends over democratic means (BF_{01} = 17.76). Taken together, the results of Study 3 provide further support that interventions that reduce affective polarization do not reduce anti-democratic attitudes. We even found some evidence that the warm elite relations intervention can increase anti-democratic attitudes.

Meta-Analyses

To provide tests of the effect of the depolarization interventions on anti-democratic attitudes with the largest possible sample size, we conducted random-effects meta-analyses (39) using the effect sizes from the three studies and from two additional pilot tests we conducted (described in more detail in the Supplementary Information on Pilot Studies 1 and 2). According to power analyses conducted with metapower (40), these tests had approximately 80% power to detect |Cohen's d| ≥ 0.06 for support for undemocratic candidates, |Cohen's d| ≥ 0.08 for support for partisan violence, and |Cohen's d| ≥ 0.07 for prioritizing partisan ends over democratic means. The differences in the power analyses result from using the observed heterogeneity which was different for the different dependent variables.

The meta-analytic results support the idea that the tested interventions reliably reduced affective polarization. Participants in the intervention conditions reported significantly lower levels of affective polarization than participants in the control conditions (Cohen's d = -0.13, SE = 0.02, z = -5.77, p < .001, 95% CI for Cohen's d = [-0.17, -0.09]). However, we did not find evidence that the tested interventions reduced anti-democratic attitudes. Specifically, the depolarization interventions did not significantly reduce support for undemocratic candidates (Cohen's d = 0.03, SE = 0.02, z = 1.28, p = .200, 95% CI for Cohen's d = [-0.01, 0.06]), nor support for partisan violence (Cohen's d = -0.01, SE = 0.03, z = -0.30, p = .767, 95% CI for Cohen's d = [-0.06, 0.05]), nor prioritizing partisan ends over democratic means (Cohen's d = -0.03, SE = 0.02, z = -1.13, p = .259, 95% CI for Cohen's d = [-0.07, 0.02]).

These results do not rule out that depolarization interventions may have small effects on anti-democratic attitudes. Because the effect of the interventions on affective polarization is Cohen's d = -0.13 and our power analysis suggested that we have 80% power to detect |Cohen's d| ≥ 0.06 - 0.08, we were powered to detect downstream effects of an approximately 2:1 ratio. This means that we cannot rule out that there are downstream effects via affective polarization with a larger ratio. However, our results suggest that depolarization interventions usually only modestly reduce affective polarization itself. Thus, even if there were downstream effects via affective polarization of a 3:1 or 4:1 ratio, the downstream effects of depolarization interventions would be very small.

Nonetheless, our best estimate is that the effect of depolarization interventions on anti-democratic is essentially null. This is partly based on the estimated effect sizes. In addition, Bayesian random-effects meta-analyses found strong evidence for a null effect of the depolarization interventions on support for undemocratic candidates (BF_{01} = 24.35), support partisan violence (BF_{01} = 24.59), and prioritizing partisan ends over democratic means (BF_{01} = 16.89). Finally, the results of instrumental variable analyses were also consistent with the conclusion that reducing affective polarization does not reduce anti-democratic attitudes (see Supplementary Table 21).

Possible Mediators and Moderators

These results beg the question, where does the causal chain from depolarization interventions to anti-democratic attitudes break? In additional exploratory analyses, we did not find evidence that any of the tested interventions reduced potential mediators for the link between affective polarization and anti-democratic attitudes (e.g., strength of inparty identification, empathy for outpartisans; see Supplementary Table 6). We also failed to find evidence for reliable moderation effects by party identity (Republican vs Democrat; see Supplementary Tables 7-13) or strength of inparty identification (see Supplementary Tables 14-16).

Discussion

Our findings call into question whether depolarization interventions developed to reduce affective polarization also reduce anti-democratic attitudes. Across three experiments, we successfully replicate previous research, finding that three depolarization interventions reliably reduced self-reported affective polarization. We also extend past work by showing that these interventions also impact behavioral indicators of affective polarization, thereby demonstrating that depolarization interventions can influence behaviors with real, monetary stakes for outpartisans. Critically, however, the depolarization interventions did not reliably reduce any of three measures of anti-democratic attitudes: support for undemocratic candidates, support for partisan violence, and prioritizing partisan ends over democratic means. Even the correlational associations between affective polarization and these anti-democratic attitudes were not reliable. Thus, we conclude that many researchers (including ourselves, e.g., 2, 22) may have substantially overestimated the effects of depolarization interventions on anti-democratic tendencies.

Our paper has several important limitations. First, we focused on estimating the effects of depolarization interventions on anti-democratic attitudes. The observed null effects do not imply that depolarization interventions cannot have effects on other important measures, such as economic, social, or romantic discrimination against outpartisans. For example, we find that depolarization interventions can increase giving in a dictator game and reduce spending in a joy-of-destruction game (see also 32, 41-42). Depolarization interventions may also have other political effects. For example, prior research has found that reducing affective polarization increases willingness to compromise (12). Another limitation of our paper is that our studies do not use probability samples and, thus, are not truly representative of the US population.

However, two of our three studies were conducted on samples that were similar to the national population on several key demographics. Another limitation is that we tested only a subset of existing depolarization interventions (although we tested interventions that involved both voters and elites). Future researchers should test additional depolarization interventions to identify what kind of interventions - if any - simultaneously reduce affective polarization and anti-democratic attitudes. Future researchers should also think carefully about what kind of affective polarization they want to measure (see also 43). It could be that the joy of destruction game is better suited to measure hate while feeling thermometers and the dictator game capture dislike. A final limitation of our paper is that we tested only a limited set of potential mediators and moderators. Future research should examine if other constructs (e.g., anti-establishment orientations, 44) moderate the effects of depolarization interventions and test which mediators interventions need to move to reduce anti-democratic attitudes.

Our findings are important because they replace an old assumption - that depolarization interventions will reduce anti-democratic attitudes (e.g., 2, 18, 21, 22, 25) - with an empirical-based default - that depolarization interventions do not reduce anti-democratic attitudes. That is, researchers and practitioners who are concerned about anti-democratic attitudes should not presume that treating affective polarization will impact those outcomes. Instead, they should see affective polarization and anti-democratic attitudes as two separate classes of outcomes that require distinct interventions.

More generally, our results suggest that research on depolarization interventions should avoid assuming downstream consequences on these interventions and instead measure these potential downstream consequences directly. That is, researchers and practitioners who are interested in interventions targeting anti-democratic attitudes such as support for undemocratic politicians, support for partisan violence, and prioritizing partisan ends over democratic means 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 anti-democratic attitudes actually exists. Future research is needed to examine whether variables that have been identified as leading to affective polarization (45-48) also affect anti-democratic attitudes. Future research should also examine whether anti-democratic attitudes may affect affective polarization. For example, testing whether interventions designed to reduce anti-democratic attitudes also reduce affective polarization would provide insights into whether there is no causal relationship between the two constructs or whether anti-democratic attitudes actually cause affective polarization. For these reasons, identifying factors that can reduce anti-democratic attitudes should be a priority for future research (e.g., 49).

Methods

Ethics Statement and Reproducibility

All studies were approved by the Institutional Review Board at Stanford University. All participants provided informed consent and were compensated for their participation. Materials, anonymized data (including descriptions of how the original files were anonymized), and analysis code for both studies are available via https://osf.io/n5u9d/. The preregistrations for Studies 2 and 3 are available via https://osf.io/a2ukg/ and https://osf.io/rmhct/.

Samples

All studies were conducted in Qualtrics. In Study 1, we collected data from 2,341 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. In Study 2, we collected data from 2,021 participants who were recruited from a large panel of previously recruited Amazon Mechanical Turk workers between March 23, 2021 and April 5, 2021. In Study 3, we collected data from 4,023 participants from Lucid between December 18, 2021 and January 2, 2022. The samples in Studies 1 and 3 were quota-matched so that they were similar to the national population on key demographics. All samples were recruited with soft quotas for participants' self-identified partisanship (including learners): 50% Democrats and 50% Republicans.

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 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 missing data. According to power analyses conducted with G*Power (50), all studies had at least 80% power to detect even small effect sizes ($|\text{Cohen's } d| \geq 0.14$ in Study 1, $|\text{Cohen's } d| \geq 0.12$ in Study 2, and $|\text{Cohen's } d| \geq 0.09$ in Study 3) and 95% power to detect $|\text{Cohen's } d| \geq 0.18$ in Study 1, $|\text{Cohen's } d| \geq 0.16$ in Study 2, and $|\text{Cohen's } d| \geq 0.11$ in Study 3. More details on sample characteristics, exclusion rules, and demographics are provided in Supplementary Tables 2-3.

Interventions

We chose interventions from the literature which we perceived as particularly promising ways of reducing affective polarization. Our focus was more on testing whether these interventions would reduce anti-democratic attitudes than on replicating the original effects as closely as possible. Therefore, our design differed in several ways from the original studies, including the control conditions, the dependent variables, and the analyses conducted.

Participants were randomly assigned to one of three (Study 1) or two (Studies 2 and 3) conditions. The first intervention we tested was a friendship intervention (12). In this condition, which we 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.”

The second intervention we tested was a misperception correction intervention. The original paper (16) tested two versions of this intervention and we selected the hypocrisy prevention intervention because it had a descriptively bigger effect size (although not significantly different). In this condition, which was included in Studies 2 and 3, 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).

The third intervention we tested was a warm elites relation intervention. This intervention is similar to the warm elite relations treatments used by prior research (15) but avoids deception. In this condition, which was included in Study 3, participants were asked to watch a video about the friendship between the Democratic politician Joe Biden and the Republican politician John McCain. The video is available at shorturl.at/jrOP6.

We used two different control conditions. In the first two studies, we used a null control. That is, in the control condition, participants moved immediately towards the section with the measures of the dependent variable. In the third study, participants watched a video about the history of neckties.

Measures

We measured affective polarization using a feeling thermometer rating for outpartisans: "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 so that higher scores indicate colder feelings

An alternative operationalization of affective polarization is the difference score of feelings toward inpartisans and outpartisans. We found that the interventions tested in Studies 1 and 2 significantly reduced the difference score (all ps ≤ .001), and the intervention tested in Study 3 reduced the difference score directionally but not significantly (p = .136). Detailed results are reported in Supplementary Table 4. We report the outpartisan feeling thermometer as our main measure in the manuscript for several reasons. First, the difference score could be decreased by (a) decreasing cold feelings toward outpartisans or (b) by increasing cold feelings toward inpartisans. Because increasing cold feelings toward inpartisans is not clearly normatively desirable, we focused on decreasing cold feelings toward outpartisans. Second, prior work suggests that feelings toward outpartisans animosity is the major source of change in affective polarization over time (51). Third, we preregistered the outpartisan feeling thermometer as our main measure in Studies 2 and 3. In Study 3, we also examined whether the intervention would impact feelings towards outpartisan voters and politicians differently (see Supplementary Table 5).

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. How many cents (if any) will you give to the [Republican/Democratic] participant?”. Withholding money in a dictator game is a behavioral measure of affective polarization used in a seminal article on the topic (1).

Spending money in a joy-of-destruction game was measured with the following item (based on 37): “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. How many cents (if any) do you want to pay to remove the [Republican/Democratic] participant’s earnings? (remember, for every 1 cent you pay, the [Republican/Democratic] participant will lose 2 cents).” Although spending money in a joy of destruction game is not a traditional measure of affective polarization, we included it because it measures hate - a stronger form of animosity - than either the outpartisans feeling thermometer or withholding money in the dictator game which capture dislike (see Supplementary Information on Correlational Statistics for the Joy of Destruction Game).

Support for undemocratic 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 35): (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 Studies 2 and 3. The items formed a reliable scale in all studies (Study 1: Cronbach's α = .90; Study 2: Cronbach's α = .88; Study 3: Cronbach's α = .91).

Support for partisan violence was measured with the following four items (36): (i) When, if ever, is it OK for [Democrats/Republicans] to send threatening and intimidating messages to [Republican/Democratic] party leaders?, (ii) When, if ever, is it OK for an ordinary [Democrat/Republican] in the public to harass an ordinary [Republican/Democrat] on the Internet, in a way that makes the target feel frightened?, (iii) How much do you feel it is justified for [Democrats/Republicans] to use violence in advancing their political goals these days?, and (iv) How much do you feel it is justified for [Democrats/Republicans] to use violence if the [Republican/Democratic] 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". The items formed a reliable scale in all studies (Study 1: Cronbach's α = .95; Study 2: Cronbach's α = .93; Study 3: Cronbach's α = .96).

Prioritizing partisan ends over democratic means 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 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 Studies 2 and 3. The items formed a reliable scale in all studies (Study 1: Cronbach's α = .84; Study 2: Cronbach's α = .85; Study 3: Cronbach's α = .91). Please note that while we refer to this variable in our preregistration as support for undemocratic practices, we think that labeling the scale prioritizing partisan ends over democratic means is more accurate. Descriptive statistics for these measures are discussed in the Supplementary Information on Descriptive Statistics. We also included additional dependent variables that are not relevant for answering the research questions of this paper. The questionnaires for all studies are available via https://osf.io/n5u9d/?view_only=bd46d6d6d32e4a43ac67130639788280.

Analysis Strategy

Our main analysis strategy (preregistered in Studies 2 and 3) was null hypothesis significance testing. For the (non-preregistered) correlational analyses, we used the Pearson correlation coefficient. For the experimental analyses, we used linear regression analyses, controlling for participants' gender, age, race, education, partisan identity, and strength of partisan identity. We used p-values from two-tailed tests as our inference criteria. In addition, we conducted a robustness check (preregistered in Study 3) using the Holm-Bonferroni adjustment for multiple hypothesis testing when we used multiple dependent variables for a construct (such as the three measures of anti-democratic attitudes).

For the non-significant effects, we conducted exploratory Bayesian analyses to estimate the strength of the evidence in favor of the null hypothesis. We conducted Bayesian ANCOVAs in JASP including the same control variables as described above. We used JASP's default settings for priors that the null hypothesis and alternative hypothesis are equally likely to be true (for robustness checks with different priors, see Supplementary Figures 1-12).

For the meta-analysis, we conducted frequentist random-effects meta-analyses using the R package metafor (39). Bayesian random-effects meta-analyses were conducted via JASP. We used seven effect sizes from the three studies and from two additional pilot tests we conducted (described in more detail in the Supplementary Information on Pilot Studies 1 and 2). Because some of these effect sizes rely on comparisons with the same control condition, we conducted robustness checks accounting for this dependency. These robustness checks provided converging results (see Supplementary Table 17).

Data Availability

The data for our studies are openly available via https://osf.io/n5u9d/.

Code Availability

The analysis scripts for our studies are openly available via https://osf.io/n5u9d/.

Acknowledgement

The authors received funding for this project from the Civic Health Project (RW), the Stanford Center on Philanthropy and Civil Society (RW), and the Institute for Policy Research, Northwestern University (JND). This work is supported under a Stanford Interdisciplinary Graduate Fellowship (JGV). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

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.

Figure Legends

Figure 1: Effects of the Friendship Intervention and the Misperception Correction Intervention on Affective Polarization and Anti-Democratic Attitudes, Study 1. For each condition and outcome, the figure shows a boxplot (left), a halved violin plot (middle), and a point cloud (right). The box of the boxplot shows the 25th percentile, the median, and the 75th percentile. The length of the whiskers is 1.5*IQR unless the minimum/maximum fall within 1.5*IQR of the quartiles. For affective polarization and withholding money in dictator game, $n_C = 835$, $n_F = 754$, $n_M = 752$; spending money in joy of destruction game: $n_C = 829$, $n_F = 751$, $n_M = 751$; support for undemocratic politicians: $n_C = 806$, $n_F = 736$, $n_M = 736$; support for partisan violence: $n_C = 812$, $n_F = 744$, $n_M = 740$; prioritizing partisan ends over democratic means: $n_C = 833$, $n_F = 751$, $n_M = 750$. Please note that the two economic games used different ranges (0-25 and 0-50) than the other dependent variables (0-100).

Figure 2: Effects of the Misperception Correction Intervention on Affective Polarization and Anti-Democratic Attitudes, Study 2. For each condition and outcome, the figure shows a boxplot (left), a halved violin plot (middle), and a point cloud (right). The box of the boxplot shows the 25th percentile, the median, and the 75th percentile. The length of the whiskers is 1.5*IQR unless the minimum/maximum fall within 1.5*IQR of the quartiles. For affective polarization, $n_C = 1016$, $n_M =1005$; support for undemocratic politicians: $n_C = 1010$, $n_M = 1002$; support for partisan violence: $n_C = 1012$, $n_M = 1004$; prioritizing partisan ends over democratic means: $n_C = 1013$, $n_M = 1005$.

Figure 3: Effects of the Warm Elite Relations Intervention on Affective Polarization and Anti-Democratic Attitudes, Study 3. For each condition and outcome, the figure shows a boxplot (left), a halved violin plot (middle), and a point cloud (right). The box of the boxplot shows the 25th percentile, the median, and the 75th percentile. The length of the whiskers is 1.5*IQR unless the minimum/maximum fall within 1.5*IQR of the quartiles. For affective polarization, n_C = 1998, n_W =2025; support for undemocratic politicians: n_C = 1962, n_W = 2006; support for partisan violence: n_C = 1993, n_W = 2017; prioritizing partisan ends over democratic means: n_C = 1990, n_W = 2016.

References

1. S. Iyengar, S. J. Westwood, Fear and loathing across party lines: New evidence on group polarization. Am. J. Pol. Sci. 59, 690-707 (2015).
2. E. J. Finkel et al., Political sectarianism in America: A poisonous cocktail of othering, aversion, and moralization. Science 370, 533-536 (2020).
3. S. Iyengar, Y. Lelkes, M. Levendusky, N. Malhotra, N., S. J. Westwood, The origins and consequences of affective polarization in the United States. Annu. Rev. Polit. Sci 22, 129-146 (2019).
4. L. Boxell, M. Gentzkow, J. Shapiro, Cross-country trends in affective polarization. National Bureau of Economic Research [Preprint] (2020). https://www.nber.org/papers/w26669
5. S. Iyengar, M. Krupenkin, The strengthening of partisan affect. Polit. Psychol. 39, 201-218 (2018).
6. Pew Research Center, Partisan antipathy: more intense, more personal (2019). https://www.pewresearch.org/politics/2019/10/10/the-partisan-landscape-and-views-of-the-parties/
7. NBC News/Wall Street Journal Survey, Study #181259 (2018). http://wsj.com/public/resources/documents/181259NBCWSJOctober2018PollFinal.pdf
8. A. I. Abramowitz, S. Webster, The rise of negative partisanship and the nationalization of US elections in the 21st century. Elect Stud 41, 12-22 (2016).
9. D. Diermeier, C. Li, Partisan affect and elite polarization. Am. Polit. Sci. Rev. 113, 277-281 (2019).
10. M. J. Hetherington, T. J. Rudolph, Why Washington Won't Work: Polarization, Political Trust, and the Governing Crisis (University of Chicago Press, 2015).

11. E. Klein, Why We're Polarized (Simon and Schuster, 2020).

12. M. S. Levendusky, Our Common Bonds: Using what Americans Share to Help Bridge the Partisan Divide. (Unpublished Manuscript, University of Pennsylvania, 2020). https://cpb-us-w2.wpmucdn.com/web.sas.upenn.edu/dist/9/244/files/2020/10/ocb_for_review.pdf

13. L. Mason, Uncivil Agreement: How Politics Became Our Identity (University of Chicago Press, 2018).

14. D. J. Ahler, G. Sood, The parties in our heads: Misperceptions about party composition and their consequences. J. Polit. 80, 964-981 (2018).

15. L. Huddy, O. Yair, Reducing affective polarization: Warm group relations or policy compromise?. Polit. Psychol. 42, 291-309 (2021).

16. J. Lees, M. Cikara, Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. Nat. Hum. Behav. 4, 279-286 (2020).

17. M. S. Levendusky, Americans, not partisans: Can priming American national identity reduce affective polarization?. J. Polit. 80, 59-70 (2018).

18. S. L. Moore-Berg, L. O. Ankori-Karlinsky, B. Hameiri, E. Bruneau, Exaggerated meta-perceptions predict intergroup hostility between American political partisans. Proc. Natl. Acad. Sci. U.S.A. 117, 14864-14872 (2020).

19. K. Ruggeri et al., The general fault in our fault lines. Nat. Hum. Behav., 1-11 (2021).

20. O. Simonsson, J. Marks, Love thy (partisan) neighbor: Brief befriending meditation reduces affective polarization. SSRN [Preprint] (2020). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3674051

21. S. Swanson, By the people: The role of local deliberative forums in combating affective political polarization. The Project on International Peace and Security (2021). https://www.wm.edu/offices/global-research/_documents/pips/selene-swanson-whitepaper

22. J. G. Voelkel, D. Ren, M. J. Brandt, Inclusion reduces political prejudice. J. Exp. Soc. Psychol. 95, 104149 (2021).

23. B. R. Warner, H. K. Horstman, C. C. Kearney, Reducing political polarization through narrative writing. J. Appl. Commun. Res. 48, 459-477 (2020).

24. M. Wojcieszak, B. R. Warner, Can interparty contact reduce affective polarization? A systematic test of different forms of intergroup contact. Polit. Commun. 37, 789-811 (2020).

25. A. Zoizner, S. R. Shenhav, Y. Fogel-Dror, T. Sheafer, Strategy news is good news: How journalistic coverage of politics reduces affective polarization. Polit. Commun., 1-20 (2020).

26. R. Hartman, Rachel, J. W. Blakey, J. Womick, C. A. Bail, E. Finkel, J. Schroeder, P. Sheeran, J. J. Van Bavel, R. Willer, K. Gray, Interventions to reduce partisan animosity [Preprint] (2022). https://psyarxiv.com/ha2tf/

27. N. Gidron, J. Adams, W. Horne, American Affective Polarization in Comparative Perspective (Cambridge University Press, 2020).

28. J. Kingzette, J. Druckman, S. Klar, Y. Krupnikov, M. Levendusky, J. B. Ryan, How Affective Polarization Undermines Support for Democratic Norms. Public. Opin. Q. 85, 663-677 (2021).

29. J. Lees, M. Cikara, Understanding and combating misperceived polarization. Philos Trans R Soc B, 376, 20200143 (2021).

30. J. McCoy, M. Sommer, Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies. Ann. Am. Acad. Pol. Soc. Sci. 681, 234–271 (2019).

31. Y. E. Orhan, The relationship between affective polarization and democratic backsliding: comparative evidence. Democratization, 29, 714-735 (2022)

32. D. E. Broockman, J. L. Kalla, S. J. Westwood, Does affective polarization undermine democratic norms or accountability? Maybe not. Open Science Framework (forthcoming). Am. J. Pol. Sci. https://osf.io/9btsq/download

33. R. E. Carlin, G. J. Love, Political competition, partisanship and interpersonal trust in electoral democracies. Br. J. Polit. Sci. 48, 115–139 (2016).

34. S. Whitt, A. B. Yanus, B. McDonald, J. Graeber, M. Setzler, G. Ballingrud, M. Kifer, Tribalism in America: behavioral experiments on affective polarization in the Trump era. Journal of Experimental Political Science, 8, 247-259 (2021).

35. M. H. Graham, M. W. Svolik, Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States. Am. Polit. Sci. Rev. 114, 392-409 (2020).

36. N. P. Kalmoe, L. Mason. Lethal mass partisanship: Prevalence, correlates, and electoral contingencies. [Preprint] (2019). https://www.dannyhayes.org/uploads/6/9/8/5/69858539/kalmoe___mason_ncapsa_2019_-_lethal_partisanship_-_final_lmedit.pdf

37. K. Abbink, A. Sadrieh, The pleasure of being nasty. Econ. Lett. 105, 306-308 (2009).

38. J. van Doorn et al., The JASP guidelines for conducting and reporting a Bayesian analysis. Psychonomic Bulletin & Review 28, 813-826 (2021).

39. W. Viechtbauer, Conducting meta-analyses in R with the metafor package. J Stat Softw 36, 1-48 (2010).

40. J. W. Griffin, Calculating statistical power for meta-analysis using metapower. Quant Method Psychol, 17, 24-39 (2021).

41. K. Gift, T. Gift, Does politics influence hiring? Evidence from a randomized experiment. Polit Behav, 37, 653-675 (2015).

42. C. McConnell, Y. Margalit, N. Malhotra, M. Levendusky, The economic consequences of partisanship in a polarized era. Am J Pol Sci, 62, 5-18 (2018).

43. E. C. Cassese, Partisan dehumanization in American politics. Polit Behav, 43, 29-50 (2021).

44. J. E. Uscinski et al., American politics in two dimensions: Partisan and ideological identities versus anti‐establishment orientations. Am J Pol Sci, 65, 877-895 (2021).

45. L. D. Bougher, The correlates of discord: identity, issue alignment, and political hostility in polarized America. Polit Behav, 39, 731-762 (2017).

46. L. Mason, A cross-cutting calm: How social sorting drives affective polarization. Public Opin Q, 80, 351-377 (2016).

47. L. A. Santos, J. G. Voelkel, R: Willer, J. Zaki, Belief in the utility of cross-partisan empathy reduces partisan animosity and facilitates political persuasion. Psychol Sci (forthcoming). https://drive.google.com/file/d/1lzJr4EMfAcVRnecS7l9BXtQ7Lb8KajY-/view

48. E. N. Simas, S. Clifford, J. H. Kirkland, How empathic concern fuels political polarization. Am Polit Sci Rev, 114, 258-269 (2020).

49. L. M. Bartels, Ethnic antagonism erodes Republicans' commitment to democracy. Proc. Natl. Acad. Sci. U.S.A. 117, 22752-22759 (2020).

50. F. Faul, E. Erdfelder, A. Buchner, A. G. Lang, Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behav. Res. Methods. 41, 1149-1160 (2009).

51. E. Groenendyk, E. Competing motives in a polarized electorate: political responsiveness, identity defensiveness, and the rise of partisan antipathy. Polit Psychol, 39, 159-171 (2018)