Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not*

David E. Broockman† Joshua L. Kalla‡ Sean J. Westwood§

First draft: December 21, 2020
This draft: September 30, 2021

Abstract

Scholars warn that affective polarization undermines democratic norms and accountability: they speculate that if citizens were less affectively polarized, they would be less likely to endorse norm violations, overlook copartisan politicians' shortcomings, oppose compromise, adopt their party's views, or misperceive economic conditions. We argue the contrary: affective polarization is not likely to influence political choices. We support this argument with four experiments which manipulate citizens' affective polarization with a trust game and trace downstream consequences, such as reactions to information about their representatives. In these experiments (total $N = 9,837$), we produce the equivalent of three decades of change in affective polarization, but find no evidence that these changes influence many political behaviors—only some general questions about interpersonal attitudes. A fifth experiment ($N = 2,504$) finds similar results with alternative manipulations of affective polarization. Our results suggest caution about assuming that reducing affective polarization would meaningfully bolster democratic norms or accountability.

*This research was approved by the IRBs at the University of California, Berkeley, Stanford University, Yale University, and Dartmouth College. Broockman acknowledges the Civic Health Project for support.

†Associate Professor, Travers Department of Political Science, University of California, Berkeley. dbroockman@berkeley.edu.

‡Assistant Professor, Department of Political Science and Department of Statistics and Data Science, Yale University. josh.kalla@yale.edu.

§Associate Professor, Department of Government, Dartmouth College. sean.j.westwood@dartmouth.edu.

Affective polarization—citizens' more negative sentiment towards opposing political parties than their own (Iyengar, Sood and Lelkes 2012)—has been growing worldwide (Gidron, Adams and Horne 2020). Research on this trend constitutes one of the most influential literatures in contemporary social science and has sown alarm across disciplines (e.g., Finkel et al. 2020).

The gravest worries the literature raises are about democracy: that affective polarization has *political* consequences, such as changing the actions citizens incentivize politicians to take, or shifting the norms to which they expect their representatives to adhere. In Table 1, we give examples of over a dozen studies that express concern about the political consequences of affective polarization. Scholars therefore urge efforts to decrease affective polarization, hoping that reducing affective polarization will have salutary effects across many political domains.

Surprisingly, however, research investigating potential political consequences of affective polarization is scarce. Existing literature has generally simply asserted that affective polarization must have downstream consequences for political decisions, with scant attention paid to the theoretical process by which this might happen and with little supporting empirical evidence. In other words, speculation is rife, but we actually "know little" (Druckman et al. 2020*b*, p. 9).

In this paper, we remedy this dearth of theory and evidence.

Theoretically, returning to the insights of foundational research on affect and decision-making, we describe why we might expect affective polarization's consequences to be usually unlikely to spill over into political judgments, but be more likely to spill over into general survey questions about interpersonal decisions.

Empirically, we address one of the main challenges faced by previous research. Previous research has had difficulty investigating the political consequences of affective polarization because of the potential for omitted variable bias: people with different levels of affective polarization likely differ in other ways, confounding comparisons between them. To probe the causal effects of affective polarization on political attitudes, we conduct a series of experiments which manipulate Americans' levels of affective polarization and then probe potential

Table 1: Example Speculation On Political Implications of Affective Polarization in Prior Work

Outcome	Quote
Electoral Accountability	• "...growth in affective polarization may weaken the role of elections in moderating [elite] polarization. ...the willingness to punish one's own party's politicians for taking an extreme position will weaken..." (Pierson and Schickler 2020, p. 50) • "we suspect that affective polarization increases support for extremist politicians, or, at least, blinds partisans to the ideological extremity of candidates from their party" (Iyengar et al. 2019, p. 142) • "[P]olarization undermines the public's ability to serve as a democratic check." (Graham and Svolik 2020, p. 407) • "...heightened polarization has made it almost impossible for partisans to abandon their party's candidates, no matter their limitations" (Iyengar and Krupenkin 2018, p. 215)
Adopting Party's Policy Attitudes	• "For the mass public, we suspect that affective polarization increases partisans' willingness to conform to their party's policy positions" (Iyengar et al. 2019, p. 142) • "If affective polarization – and, most importantly, partisan animus – is associated with greater responsiveness to party cues, then elite behaviours could have tremendous capacity to change mass response to the pandemic" (Druckman et al. 2020*a*, p. 9).
Legislative Bipartisanship	• "Negative views of the opposing party among voters, in turn, encourage political elites to adopt a confrontational approach to governing." (Abramowitz and Webster 2016, p. 22) • "holding opposing partisans in contempt ...precludes innovative cross-party solutions and mutually beneficial compromises" (Finkel et al. 2020, p.533) • "...affective polarization makes governance more difficult." (Levendusky 2018, p. 59)
Democratic Norms	• "Understanding affective polarization is important because it erodes democratic norms and institutions." (Gidron, Adams and Horne 2020, p. 3) • "The cumulative effect of severe polarization as we have defined it here is a deterioration in the quality of democracy, leading to backsliding, illiberalism, and in some cases reversion to autocracy." (McCoy and Somer 2019, p. 258) • "As affective polarization increases, partisans may become more likely to ignore democratic norms..." (Kingzette et al. 2021, p. 2) • "affective polarization can have grave ramifications...Partisanship appears to now compromise the norms and standards we apply to our elected representatives, and even leads partisans to call into question the legitimacy of election results" (Iyengar et al. 2019, p. 143)
Condition Perceptions	• "These findings underscore the challenges that affective polarization poses for governance. In a polarized America, citizens may be willing to tolerate poor economic performance from their own party, or fail to reward the other side for apparently good economic stewardship, winnowing further already weak hopes that the public will be responsive to government action." (Freeder 2020, p. 28) • "[T]he increased level of affective polarization poses considerable challenges to the democratic process. Partisan bias in perceptions of economic conditions means that voters will fail to credit opposing-party incumbents when the economy grows under their stewardship and fail to penalize in-party incumbents whose economic performance is suspect." (Iyengar, Sood and Lelkes 2012, p. 428)

downstream consequences, such as consequences for how they react to information about their representatives. We manipulate affective polarization using a trust game in which respondents to a

survey believe they are playing with an outpartisan. The putative outpartisan player either is highly cooperative, resulting in the respondent earning a bonus, or completely uncooperative, resulting in no bonus for the respondent.[1] This manipulation produces very large effects on affective polarization as measured in the same manner as the literature; the exogenous differences in affective polarization we create are similar in size to the three decades of increases in affective polarization in the United States documented by Iyengar, Sood and Lelkes (2012).

Deploying this paradigm across four surveys and $N = 9,837$ respondents, we investigate the causal effects of affective polarization on a variety of downstream outcomes. First, we show that increasing affective polarization has large downstream effects on general *interpersonal* questions—general questions about other people that do not involve political candidates, institutions, or issues, and that do not introduce other relevant dimensions; e.g., generally stating discomfort with having outpartisan friends. We then investigate effects in five political domains identified in prior literature (see Table 1): electoral accountability (measured by both levels of party loyalty and how individuals react to information about their actual representatives), adopting one's party's policy positions, support for legislative bipartisanship, support for democratic norms, and perceptions of objective conditions. Our results run contrary to nearly all predictions in the literature: in these political domains, our estimates of the causal effects of reducing affective polarization are consistently null.

We also conduct a fifth survey ($N = 2,504$) as a robustness check to examine the downstream effects of three alternative and conceptually distinct approaches for manipulating affective polarization. These approaches also successfully manipulated affective polarization, but we continue to find null downstream effects on political outcomes.

Last, we conduct a simple exercise that suggests that most of the correlational relationship between affective polarization and outcomes on the general *interpersonal* questions appears to be causal, but that these correlational relationships in *political* domains appear driven by omitted

[1] A paper about linkages between partisan and racial affect developed this approach (Westwood and Peterson 2020).

variable bias.

In concluding, we discuss potential alternative explanations, remaining limitations, and potential extensions of our work.

Theoretical Framework

In this paper, we define affective polarization as traditionally conceptualized, the difference in affect one has for one's own party minus the outparty (Iyengar, Sood and Lelkes 2012).[2] In turn, affect is "a physiological state that is experienced as either pleasant (positive affect) or unpleasant (negative affect)" (Ottati and Wyer 1993, p. 297)—i.e., it refers to general feelings of positivity or negativity towards an attitude object, such as a party. Prominent theories of affect conceptualize it as representing an "overall summary evaluation" (Ottati and Wyer 1993, p. 302), collapsing judgments about many dimensions of an attitude object to a *single dimension* of overall liking or disliking. Affect is often used as a heuristic to simplify a complex judgment into a simpler question: do I like this attitude object (Slovic et al. 2002)?

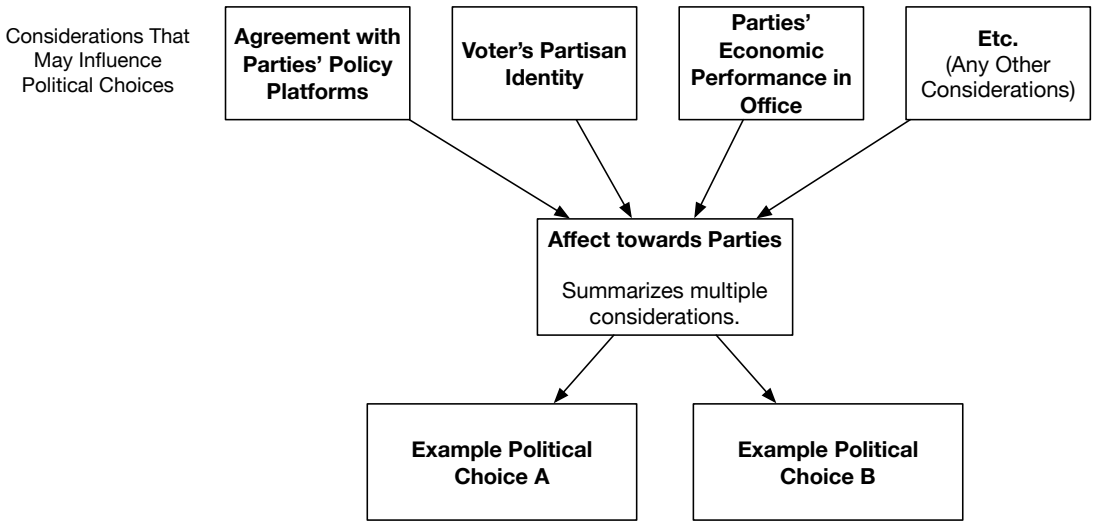
The literature's speculation about the downstream consequences of affective polarization for political judgments could therefore be consistent with a model like that shown in Figure 1a. Under this model, a number of considerations influence individuals' affect towards the parties (we are agnostic with respect to continuing debates about which). Then, when making political choices, such as whether to re-elect their member of Congress, they simply rely on their affect towards the parties as a heuristic.

We explicate two reasons why affective polarization should not so readily influence political

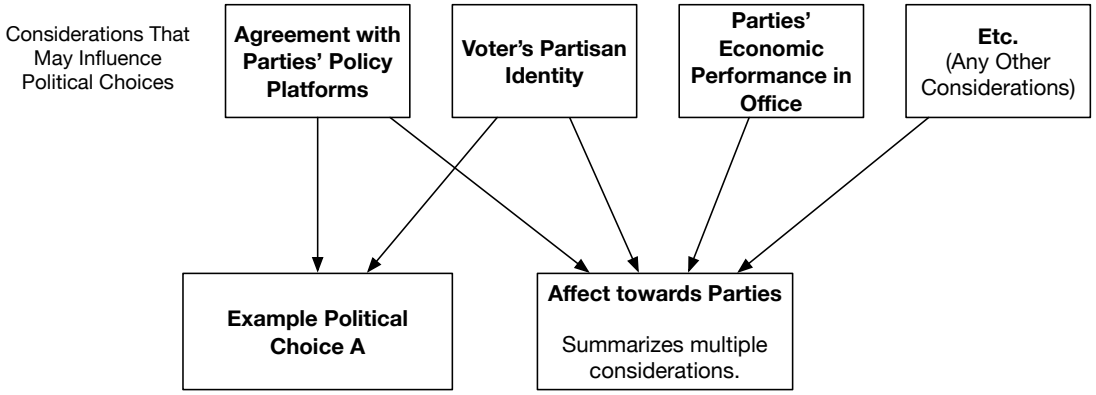
[2]This definition is also consistent with Iyengar and Westwood (2015, p. 690)'s, the "divergence in affect toward the in and out parties." Some research has considered affective polarization in the context of broader constructs, such as "social polarization" (Mason 2014), "partisan prejudice" (Lelkes and Westwood 2017), or partisan identity. When we refer to affective polarization, we refer only to the original definition and not these broader constructs. Additionally, affective polarization is often conceived as being specific to affect towards "rank-and-file" partisans, i.e., citizens (Iyengar, Sood and Lelkes 2012). However, Druckman and Levendusky (2019) note that affect is even more polarized towards outpartisan elites and that many survey items measuring affective polarization capture antipathy towards both citizens and elites. Cognizant of this, our manipulation checks distinguish between affect towards citizens and elites, and we find increases in affective polarization towards both (Figure 2).

Figure 1: Observational Equivalence Between Literature's Speculation and Alternative Hypotheses

(a) Literature's Speculation: Affective Polarization Affects Political Choices



(b) Alternative: Other Considerations Influence Both Affect, Other Political Choices (Example Choice A)



judgments.

First, research on heuristics does show that individuals often reduce complex questions—e.g., about the current unemployment rate—to simpler, related questions that are easier for them to answer—e.g., whether their closest friends have good jobs. This process is known as attribute substitution (Kahneman and Frederick 2002). However, individuals seek to substitute in replacement judgments that are both *as easy to answer* and *as relevant to the original judgment* as possible. When forming judgments about objects related to but distinct from parties, such as

candidates or norms, there are other easy-to-access attributes (and competing heuristics) that are usually more relevant than one's affect towards a party. We would expect affective polarization to principally influence those judgments where individuals do not have other, more relevant considerations they would more readily call to mind.

For example, when deciding whether to vote for a politician, voters might bring to mind a number of attributes other than their affect towards that politician's party, including their affect towards that politician, the last thing they remember hearing about that politician, their perceptions of their policy positions, the state of the economy, and their own partisan identity. Indeed, "evaluations of an object might not be based on affective reactions at all" (Ottati and Wyer 1993, p. 298) when individuals have more relevant dimensions to rely on. Even if they do rely on affect, we would expect individuals to rely first on affect towards the attitude object (e.g., the politician), and on partisan affect only if more relevant dimensions are unavailable. In other words, partisan affect could function as a fallback criterion of judgment, but there are many other, more relevant criteria we would expect voters to be able to rely on first. Consistent with this expectation, evidence has long found that citizens distinguish between outcomes that are relevant for evaluating particular politicians and those that are not (e.g., Brody and Sniderman 1977).

As shown in Figure 1b, this could produce a circumstance where political choices could be influenced by the same factors (e.g., partisan identity) that influence partisan affect, without partisan affect having any causal effect on political judgments. One could therefore imagine partisan affect manifesting in stylized survey questions about whether one would be unhappy having an outpartisan neighbor (as there are no other dimensions of judgment available) but not evaluations of an *actual* outpartisan neighbor or an *actual* outpartisan politician, when other information *is* available.[3]

[3]This does not imply that there would be no partisan bias in such encounters. For example, partisan identity—a separate construct from affect—may still produce bias against or stereotyping of the neighbor if, has been found for other social identities even in the case of groups with friendly relations, that people prefer to associate with ingroup rather than outgroup members.

A second reason to doubt that individuals make different choices due to affective polarization is that individuals must give up other things they value in order to do so—that is, *it is costly for voters to allow affective polarization to influence their judgments*. To appreciate why this is the case, note that for affective polarization to influence people's choices, it must lead them to make *different choices than they would have made if their level of affective polarization were different*. For example, in Table 2a, we first show an example where affective polarization does *not* influence a voters' judgment: they agree with an inparty candidate more on issues, and so they vote for the inparty candidate regardless of their affective polarization.[4] By contrast, in Table 2b, affective polarization *does* affect a voters' judgment. In this second scenario, a voter agrees with the outparty candidate on more issues and would vote for them were they not affectively polarized. However, if that same voter were affectively polarized, holding all else constant, affective polarization has an impact: the voter would vote for the inparty candidate instead. But notice that to change their vote, the voter must give something up: they must surrender the opportunity to vote for the candidate they agree with on more issues.

This illustrates a core principle in economics (Becker 1957): if our tastes for one dimension cause us to change our choice from A to B,[5] we must pay the cost of giving up whatever it is we liked about A on other dimensions. Put differently, presuming people have reasons to make the choices they do, in order for them to make a *different* choice, they must decide to surrender whatever benefits made their *original* choice attractive. The existence of such a trade-off is a necessary condition for affective polarization to cause people to make different choices than they would otherwise make. However, we doubt that voters would make such trades.

This second argument would also predict that affective polarization may have some effect on

[4]This illustration merely uses issue positions as an example of a dimension of candidates voters may care about. So long as voters have *any* decision-making criteria (e.g., preferences for incumbents, more qualified candidates, candidates with certain demographics, etc.), then in order for affective polarization to have an effect on judgment, voters must be willing to sacrifice their preferences on these other dimensions. This illustration could also be expanded to include the decision to abstain from voting.

[5]If a taste never changes which choice we make, that taste has no impact.

Table 2: Illustration: Allowing Affective Polarization to Affect Judgments Requires Voters to Make Trade-Offs

(a) Scenario 1: Affective Polarization Does *Not* Affect Judgment

	Inparty Candidate	Outparty Candidate
Voter Agrees with Candidate on More Issues	✓	
Votes for Candidate if Voter's Affective Polarization Low	✓	
Votes for Candidate if Voter's Affective Polarization High	✓	
Causal Effect of Affective Polarization	**None**, because voter would make same decision regardless of their affective polarization.	

(b) Scenario 2: Affective Polarization *Does* Affect Judgment

	Inparty Candidate	Outparty Candidate
Voter Agrees with Candidate on More Issues		✓
Votes for Candidate if Voter's Affective Polarization Low		✓
Votes for Candidate if Voter's Affective Polarization High	✓	
Causal Effect of Affective Polarization	**Changes** voter's vote, **but** voter must give up voting for the candidate who agreed with them on more issues.	

answers to general questions about whether one would be happy to interact with outpartisans, as there exist limited trade-offs when answering such an abstract question. However, affective polarization may have less of an effect on interactions with *specific* outpartisans for whom respondents have more information because trade-offs would exist. Consistent with this notion, work shows that adding additional information to social choice scenarios that prompts a trade off in the minds of a respondent does in fact reduce their reliance on partisanship (Klar and Krupnikov 2016).

Of course, it remains an empirical question as to whether or not citizens in fact do rely on partisan affect when forming judgments in which partisanship is only one possible criteria, and whether they would be willing to make the trade-offs necessary to do so. The goal of this theoretical

discussion is not to argue as persuasively as possible that affective polarization could not have an impact on downstream choices. Instead, we see the primary purpose of our theoretical framework as to point out the likely necessary conditions for it to do so and why these conditions are not obviously met.

However, the empirical evidence supporting the view that affective polarization has significant downstream consequences is also murky.

First, despite many scholars asserting that the causal effects of affective polarization on political outcomes are well-known (see Table 1), there is surprisingly little empirical research exploring the role of affective polarization in explicitly political contexts. Many reviews of the literature explicitly note this evidentiary lacuna (e.g., Iyengar et al. 2019; Druckman et al. 2020*b*).

Second, observational relationships between affective polarization and other outcomes are difficult to interpret: the hypotheses that affective polarization affects political judgments and that it merely correlates with political judgments are observationally equivalent in observational data. As shown in Figure 1b, insofar as affect serves to summarize other considerations, these other considerations may influence both affective polarization and political choices, leading the two to correlate even if affective polarization has no effect on political choices—a classic case of omitted variable bias. (Later in the paper, we demonstrate that exactly this appears to be occurring.) Likewise, correlated time trends between changes in mass behavior and increases in affective polarization confound any potential effects of affective polarization with other changes.

Nevertheless, understanding whether affective polarization impacts political judgments has important implications for efforts to reverse negative trends in democracies: would decreasing affective polarization bolster accountability and support for democratic norms, or does affective polarization correlate with these outcomes merely because it correlates with other more consequential dimensions? To answer this question, researchers must locate or produce an exogenous source of variation in affective polarization and then trace its impact on these downstream outcomes of interest. We do just that.

Data and Research Design

We trace the downstream impacts of affective polarization on a broad variety of political outcomes across five surveys with a total sample size of $N = 12,341$. Our surveys deploy a two-stage process in which we first randomly assign a task which changes respondents' level of affective polarization and then present a number of survey items potentially downstream of affective polarization. This design facilitates a causal analysis of the effect of affective polarization on a number of constructs.

Surveys 1-4

Our primary data is drawn from four surveys conducted in 2019 and 2020 using the online survey vendor Dynata. In order to be eligible, participants needed to provide their informed consent, identify as a Democrat or a Republican (including leaners), pass a pre-treatment attention check, and demonstrate understanding of the trust game (described below). All these criteria were assessed prior to random assignment. Our surveys resulted in a total of $N = 9,837$ completed responses. The four surveys were conducted in October 2019 (Survey 1 $N = 1,684$), December 2019 (Survey 2 $N = 2,499$), December 2019 (Survey 3 $N = 3,519$), and December 2020 (Survey 4 $N = 2,135$). The studies were all pre-registered.[6]

Appendix Table B2 presents the demographics of the survey samples relative to the 2019 Cooperative Congressional Election Study. We find that our survey samples are broadly similar to the demographics of the 2019 CCES, with the largest difference being that our sample is slightly more educated, potentially because we require our respondents to successfully complete an attention check and understand the trust game. Survey item wordings appear in Appendix C. Although all surveys contained the trust game before the outcome measures, different surveys contained different outcome measures. Appendix Table B1 lists which items appeared on which

[6]Pre-analysis plans: `https://osf.io/kde27/?view_only=9b450ae09e234f0ba3f7193432170530`, `https://osf.io/7ve49/?view_only=6d36403f548c43a6817e24c079d91d30`, and `https://osf.io/2ysp7/?view_only=1903af75960a4683a4238f835d5eff53`.

surveys and the order in which they appeared. Appendix Section C.5 discusses survey attrition.

Inducing Variation in Affective Polarization with a Trust Game

After collecting demographics, surveys 1-4 prompted participants to engage in a task from Westwood and Peterson (2020) which manipulates affective polarization.

This task uses a scripted set of allocations in a modified trust game (Berg, Dickhaut and McCabe 1995). In typical trust games, there are two players. Player 1 receives a cash allocation and is instructed to give "some, all, or none" of the money to Player 2. The player is also told that the researchers will *triple* any amount Player 1 gives to Player 2. Furthermore, Player 2 can return some, all, or none of the money back to Player 1. Therefore, the more Player 1 expects reciprocity from Player 2, the more money they should allocate to Player 2 in anticipation they will receive a larger sum in return, and the better off Player 2 will be.

Our studies deploy a version of this framework we alter in several ways. First, we always make participants take the role of Player 2. This means they always first observe an allocation another player makes to them. Second, across three consecutive rounds of game play, participants are told they are interacting with three other respondents of the opposite political party who have each been allocated \$10.[7] However, they are in fact interacting with computerized opponents who offer pre-determined allocations. Participants randomized to the Positive Experience condition receive allocations from Player 1 of \$8, \$7 and \$8 (tripled to \$24, \$21 and \$24) respectively across the three rounds of the game. However, those in the Negative Experience condition receive \$0 allocations from Player 1 in all three rounds. Following each round, in both conditions, participants are told the other player said that they made their allocation due to the player's own partisanship. At the end of the game, participants see a summary of the results for all rounds.

In the survey, participants read instructions, saw three example rounds to ensure they understood the game, and completed two comprehension questions.[8] They then participated in the

[7]The opponent's other profile attributes (gender, age, and income) were randomly assigned in each round of play.

[8]If respondents missed a comprehension question, participants were given the answer and asked the questions again. Those failing the questions three times were removed from the survey prior to random assignment.

trust game, with gameplay unfolding in a manner consistent with their random assignment. All game instructions are given in Appendix C.

Manipulation Checks

This manipulation produced large changes in affective polarization in all four of our surveys, as assessed by the main measure of affective polarization employed in the literature: the difference in feeling thermometers for citizens of the in- minus out-party. In particular, participants assigned to the Positive Experience condition rated outpartisan citizens and outpartisan politicians and elected officials much more positively than those in the Negative Experience condition.[9] (Evidence presented in Appendix D indicates that the difference between the conditions is driven by decreases in affective polarization in the Positive Experience condition.) For example, in Survey 1, the average feeling thermometer difference for in- minus out-party citizens in the Negative Experience group was 36.3, but in the Positive Experience group it shrunk to 23.8. Across our four surveys, the average difference in affective polarization towards other citizens across the Positive and Negative Experience conditions is 14.3 degrees, with most of this coming from changes in affect towards outpartisans (15.5 degrees). The effect on affective polarization towards political elites (measured only in Survey 4) is 9.8 degrees. These differences are approximately equivalent to 'rewinding' over three decades of increasing affective polarization in the United States.[10]

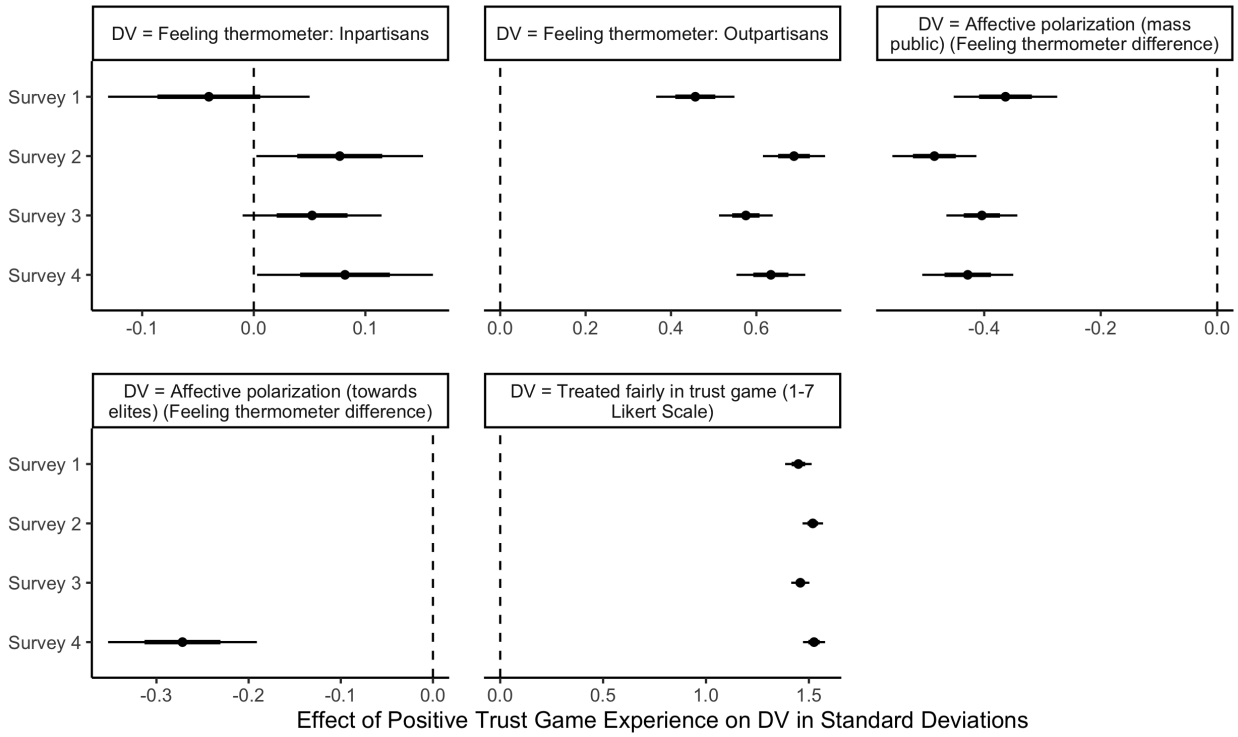
The top row of Figure 2 visualizes these effects on standardized versions of variables we asked as part of a manipulation check. As with the remaining Figures in the paper, the coefficients estimate the effect of the Positive Game Experience condition relative to the Negative Game Experience condition. Mirroring Iyengar, Sood and Lelkes's (2012) findings regarding changes

[9]These feeling thermometer items ask about "People who are Democrats/Republicans" and "Democratic/Republican Politicians and Elected Officials" in general, and do *not* reference the individual players in the game.

[10]The total increase in affective polarization Iyengar, Sood and Lelkes (2012, Table A1) observe from 1978 to 2008 is 15.35 degrees for Democrats and 10.16 degrees for Republicans, for an average of 12.76, similar to our observed effects of 14.3 and 9.8 degrees in affective polarization towards voters and elites, respectively.

over time in the electorate, we see small changes in sentiments towards inpartisans, but large increases in sentiments towards outpartisans as a result of the Positive Game Experience condition. Importantly, the intervention also meaningfully affected affective polarization towards "politicians and candidates" in general. Finally, unsurprisingly, we show that respondents were much more likely to say they were treated fairly in the trust game if in the Positive Experience condition.

Figure 2: Effect of Positive Trust Game Experience on Manipulation Checks



Notes: Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Table A2.

Results: Downstream Consequences of Affective Polarization

We next exploit these experimentally-induced reductions in affective polarization to analyze its downstream consequences across a variety of outcomes.

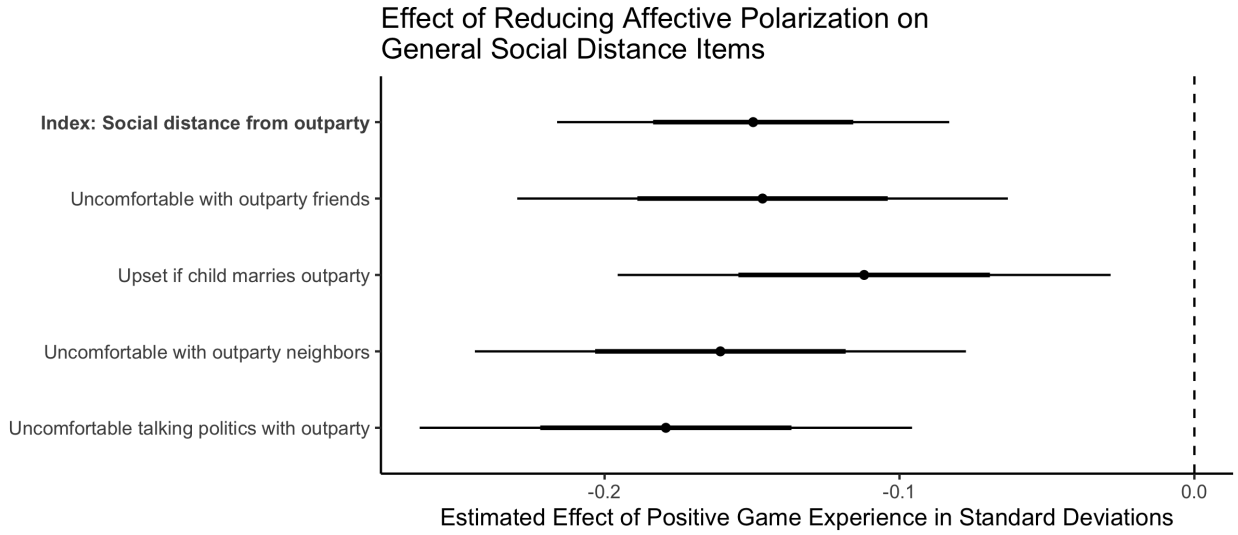
Apolitical, Interpersonal Items: General Social Distance Measures

We first examine the causal effects of affective polarization in an apolitical domain where prior research has also documented its effects: general social distance.

Figure 3 reports the effects of the treatment on a set of general social distance measures, most from Iyengar, Sood and Lelkes (2012) and Druckman and Levendusky (2019). These items probe generalized discomfort with social interactions or social ties with outparty members (the items do *not* refer to the particular partisans in the trust game). We asked these items on Survey 4.

As with the all the remaining Figures in the paper, we orient the dependent variables such that negative coefficients would correspond with the literature's prevailing predictions about what the impact would be of reducing affective polarization. Therefore, in this case, the negative estimates capture decreased discomfort with social interactions with outparty members.

Figure 3: Effect of Increasing Affective Polarization on Social Distance Items (Survey 4)



Notes: Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Table A3.

As Figure 3 shows, we observe large and statistically significant decreases on all the social distance items as a result of the positive trust game condition. The first coefficient shows an index

we pre-registered of all four of these items; the effect on this index is also significant. The treatment strongly affected all of the items we drew from previous literature, including expressing discomfort with having outpartisan friends and displeasure with one's child marrying an outpartisan.[11]

Electoral Accountability

We next turn to evaluating affective polarization's downstream consequences in political domains. Perhaps the concern scholars most commonly articulate about the political consequences of affective polarization is that it encourages voters to loyally vote for their party's candidates, no matter their shortcomings. Many scholars express some version of this concern (see Table 1), such as concern that affective polarization "has made it almost impossible for partisans to abandon their party's candidates, no matter their limitations" (Iyengar and Krupenkin 2018, p. 215).

Following from Little, Schnakenberg and Turner's (2020) framework, we consider two possible ways this concern might manifest: *divergence*, in which affective polarization simply increases party loyalty; and *desensitization*, in which affective polarization weakens voter's responses to information about incumbent's actions. If *divergence* holds, we would expect that as an individual's level of affective polarization decreases, they would become relatively more willing to support outparty politicians. If *desensitization* holds, we would expect that as individual's level of affective polarization decreases, they would become more willing to hold incumbent politicians accountable for their actions (e.g., more willing to electorally punish a copartisan MC for taking issue positions they disagree with or reward outpartisan MCs for positions they agree with).

We sought to assess these consequences of affective polarization with an experiment with an unusual degree of realism. In particular, in Surveys 1 and 3, we first asked respondents for their full zip code so that we could determine their actual Congressional district and, in turn, their actual representative in Congress (Member of Congress, MC). We next asked them whether they

[11] However, see Klar, Krupnikov and Ryan (2018) for an important critique of the marriage item.

personally supported or opposed a series of bills that Congress had previously voted on (see full text in Appendix C). Respondents then played the trust game and were randomized to have a positive or negative experience.

Finally, we showed respondents how their MC actually voted on several bills, randomizing whether we showed respondents *congruent* votes—up to 3 votes where they *agreed* with how their MC voted—or *incongruent* votes—up to 3 votes where they *disagreed* with how their MC voted. We determined the congruent and incongruent votes by comparing respondent's pre-treatment stated preferences on each issue to how their MC actually voted. That is, respondents were randomized to learn about votes their representative had cast that were either in line with respondents' pre-treatment issue positions or out of step with respondents' pre-treatment positions. In Survey 3, we also included a pure control group where respondents received no information about how their MC voted.

We concluded by asking a series of questions capturing vote intention in the next Congressional election and approval of their MC. As we pre-registered, we combine these into a MC Approval Index,[12] our primary outcome. This index is standardized to have mean zero and standard deviation one in each survey.

Respondents were told their MC's party when they learned about their votes and the MC's party affiliation was again given in the survey question asking for their vote intentions and approval. See Online Appendix C.2 for the full items and see Table B1 for the order in which all of the questions appeared.

This design produced the $2x3$ factorial design shown in Table 3. The first factor, shown at the top, is the affective polarization manipulation: whether respondents had a Positive or Negative Game Experience, which manipulated their levels of affective polarization. The second factor, shown at left, is the MC vote manipulation: whether respondents were shown votes their MC cast

[12]As we pre-registered, we standardize these three questions and compute a additive index of them. In Study 1, $\alpha = 0.93$; in Study 3, $\alpha = 0.83$.

which were either congruent or incongruent with their own views, or no votes. Respondents are therefore assigned to one of the six cells shown in the center of Table 3.

Table 3: 2x3 Design: MC Accountability Experiment

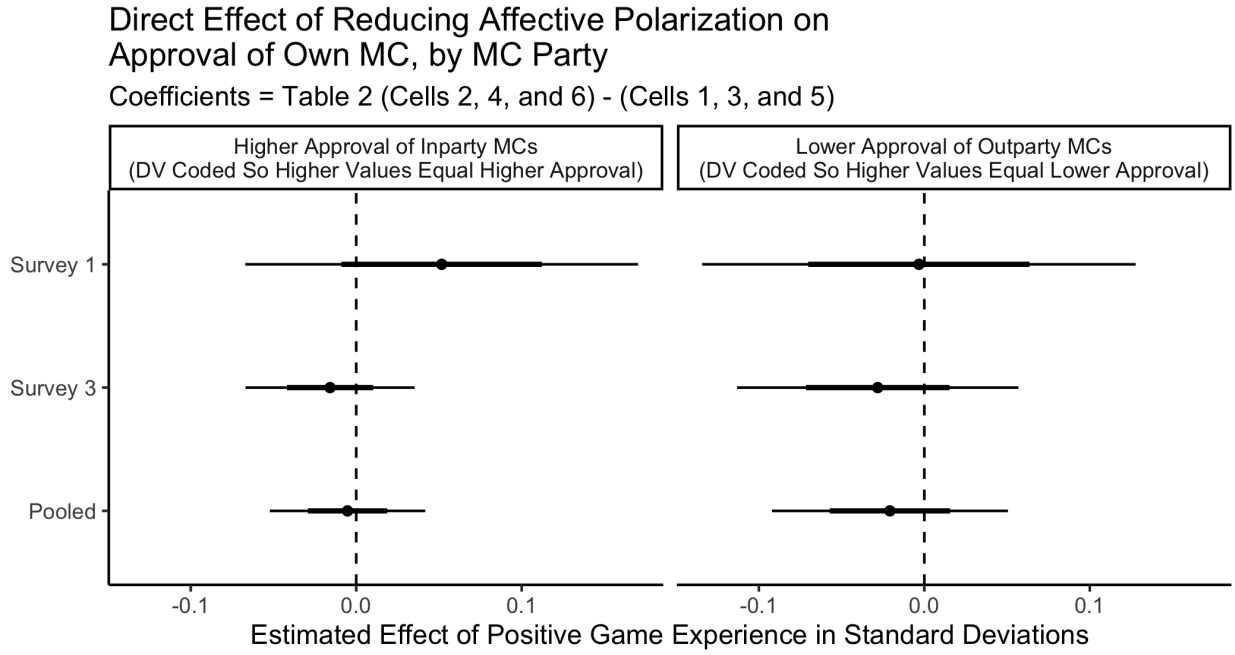
		Affective Polarization Manipulation		
		Column 1: Negative Game Experience (Higher Affective Polarization)	Column 2: Positive Game Experience (Lower Affective Polarization)	*Difference Between Columns 1 and 2*
Member of Congress Vote Manip-ulation	Row 1: Show Congruent Votes (Votes Respondent Agrees With)	Cell 1	Cell 2	*Effect of Reducing Affective Polarization, Given Respondent Shown Congruent Votes*
	Row 2: Show Incongruent Votes (Votes Respondent Disagrees With)	Cell 3	Cell 4	*Effect of Reducing Affective Polarization, Given Respondent Shown Incongruent Votes*
	Row 3 (Survey 3 Only): Show No Votes (Control)	Cell 5	Cell 6	*Effect of Reducing Affective Polarization, Given Respondent Shown No Votes*
	Difference Between Rows 1 and 2	*Effect of Showing Incongruent vs. Congruent Votes, Under Higher Affective Polarization*	*Effect of Showing Incongruent vs. Congruent Votes, Under Lower Affective Polarization*	
	Difference Between Rows 1 and 3	*Effect of Showing Congruent vs. No Votes, Under Higher Affective Polarization*	*Effect of Showing Congruent vs. No Votes, Under Lower Affective Polarization*	
	Difference Between Rows 3 and 2	*Effect of Showing No vs. Incongruent Votes, Under Higher Affective Polarization*	*Effect of Showing No vs. Incongruent Votes, Under Lower Affective Polarization*	

Divergence: Does Affective Polarization Increase Party Loyalty?

If voters were less affectively polarized, would they be less loyal to their own party, and more willing to vote for the outparty? To test this, for the moment we set aside and hold constant the MC vote manipulation and focus on the main effects of the affective polarization manipulation. In particular, we compare the MC Approval Index in the Positive and Negative Game Experience conditions; i.e., in Table 3, we compare Cells 2, 4, and 6 to Cells 1, 3, and 5.

Figure 4 shows the results. As in all our Figures, we code the dependent variables such that prevailing predictions in the literature about the impact of reducing affective polarization correspond with negative coefficients. In each panel, we show the results for both Survey 1 and Survey 3 separately, followed by the pooled results of both coefficients.

Figure 4: Testing for Divergence: Does Reducing Affective Polarization Reduce Approval of Copartisan MCs or Increase Approval of Outpartisan MCs?



Notes: Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Table A4.

The left panel subsets to respondents with copartisan MCs and tests whether respondents approve less of their copartisan MCs when their level of affective polarization is decreased (e.g., due to decreased party loyalty). If this were the case, we would expect to see negative coefficients, i.e., that reducing affective polarization decreased loyalty to inparty MCs. We do not. The right panel subsets to respondents with outpartisan MCs and shows that decreasing participants' level of affective polarization does not make them disapprove of their outpartisan MCs any less; i.e., we do not see respondents showing decreased hostility to outparty MCs when they are less affectively polarized. The point estimates in both studies in both panels are essentially zero, with tight confidence intervals.

Desensitization: Does Affective Polarization Weaken Reactions to Information about Politicians' Actions?

Another, more subtle possibility is that affective polarization weakens voter's reactions to politicians' actions in office. Following Little, Schnakenberg and Turner (2020), we call this *desensitization*, referring to the possibility that affective polarization causes people to be less sensitive to how politicians behave in office.[13]

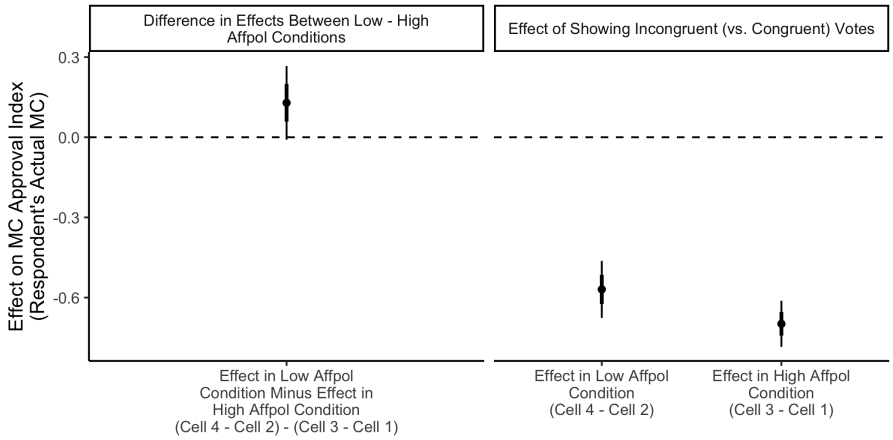
To test for this, we first separately estimate the effects of showing respondents incongruent votes their MC cast (i.e., the "Differences Between" rows at the bottom of Table 3) broken down by affective polarization condition. The desensitization hypothesis would predict that the effects of showing respondents information about how their MC voted would be larger once respondents' level of affective polarization is reduced; i.e., in the Positive Game Experience condition.

Consider first the results in Figure 5a, which shows results from Survey 1. The left panel of Figure 5a plots the effects of showing incongruent (instead of congruent) votes on the MC Approval Index, separated by affective polarization condition. The first coefficient, above "Effect in High Affpol Condition" shows that, given a respondent has been randomly assigned to the High

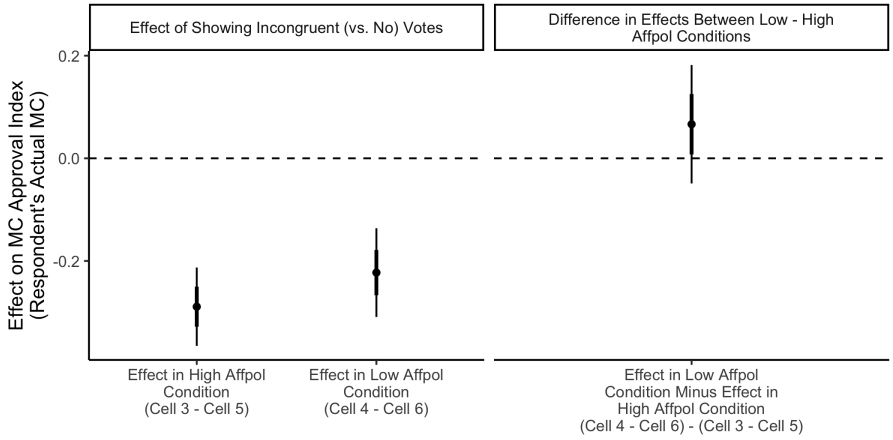
[13] An alternative possibility not considered in the literature is that affective polarization could *strengthen* the relationship between incumbent actions and voter behavior. For example, it could increase voters' willingness to punish out-of-step in-party politicians, perhaps for perceived disloyalty.

Figure 5: Testing for Desensitization: Does Increasing Affective Polarization Reduce Voter's Sensitivity to Information about their Member of Congress' Votes?

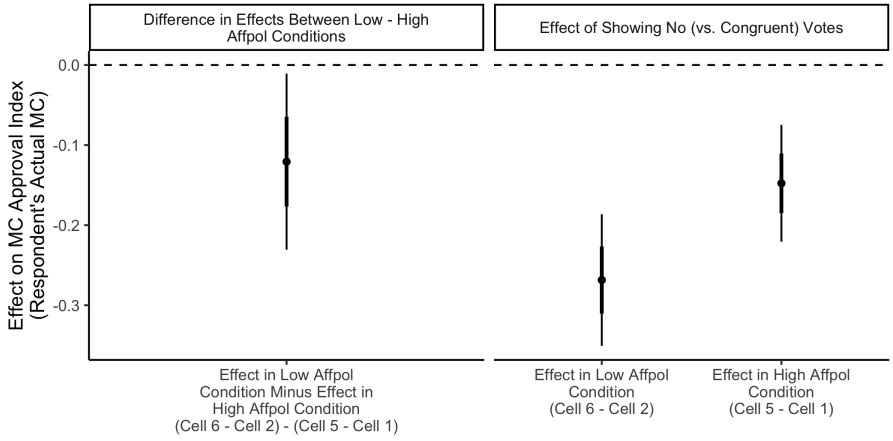
(a) Survey 1: Effect of Showing Incongruent (vs. Congruent) Votes



(b) Survey 3: Effect of Showing Incongruent (vs. Not Showing Votes) Votes



(c) Survey 3: Effect of Showing Congruent (vs. Not Showing) Votes



Notes: Cells refer to cells in Table 3. See Figure A1 for condition means and Figure A2 for robustness (e.g., effects split by whether MC is copartisan or outpartisan). Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Tables A5 - A7.

Affective Polarization condition, showing them incongruent instead of congruent votes their MC cast causes them to approve of their MC much less ($d = -0.70$). (This coefficient compares Cells 3 and 1 in Table 3.)

The key test for desensitization is how the effect of showing incongruent instead of congruent votes differs in the experimental condition with lower affective polarization. If the literature's speculation is right, we should see respondents be *more* sensitive to information about how their MC voted when they are *less* affectively polarized. However, the second coefficient shown in Figure 5a if anything indicates the opposite, as the effect of showing incongruent votes is actually slightly *smaller* ($d = -0.57$) when respondents' level of affective polarization is decreased. The difference in these two effects between the high and low affective polarization condition is $d = 0.13$; respondents react if anything slightly less negatively to the incongruent votes when their level of affective polarization is decreased.

In Survey 3 we replicated these results with a larger sample and included a pure control that was shown no information about how their Member of Congress voted.

The results are generally similar. Figure 5b shows the effects of showing respondents incongruent votes their MC cast as compared to not showing them any votes their MC cast. As in Survey 1, we see similar effects regardless of whether respondent's level of affective polarization was reduced or increased; and, if anything, those in the decreased affective polarization condition again punished their Member of Congress *less* for casting incongruent votes.

When examining the effects of casting congruent votes, we do see some sign that respondents rewarded their MCs more for casting congruent votes when in the increased affective polarization condition, a borderline statistically significant result ($p = 0.03$). However, there are several reasons to be skeptical of this result: 1) it goes in the opposite direction as the other two results, and pooling the three results yields a null result; 2) as we discuss at the end of the paper, adjusting this result for multiple comparisons renders it insignificant; and 3) while insignificant in both cases, the point estimate is similar in size for participants with copartisan and outpartisan MCs

(see Figure A2), inconsistent with what the standard view of how affective polarization would change reactions to congruent votes would hold (i.e., that the effect of reducing affective polarization on how people react to information about congruent votes, if real, would be driven by affective polarization leading respondents to more readily reward outpartisan politicians for congruent votes).

In summary, randomly inducing a reduction in affective polarization does not appear to encourage voters to engage in additional electoral accountability; if anything, on net, the point estimates on average indicate they engage in less of it. We also show in Figure A2 that these results hold if we look separately by whether the Member of Congress is a copartisan or outpartisan (indicating the results are not due to ceiling or floor effects), as well as if we focus on just a survey item capturing vote intentions and not the MC Approval Index.[14]

Adopting One's Party's Positions

Several scholars have also speculated that "affective polarization" might be "...associated with greater responsiveness to party cues" (Druckman et al. 2020*b*, p. 9). This could undermine accountability if legislators could avoid responsibility for their actions by more easily persuading affectively polarized copartisan citizens of the merits of their actions. We examine this possibility with a series of additional dependent variables we asked after the MC Accountability experiments described above. In particular, after playing the trust game, learning their MC's legislative voting record, and answering questions about favorability towards their MC, we also asked respondents to again answer questions about their own issue preferences on the issues on which we showed them their MC's votes.

Are respondents less likely to adopt their party's positions on issues when their level of affective polarization is decreased? To test this, we examine whether respondents who were assigned the positive trust game experience (low affective polarization) are less likely to express post-treatment

[14]Figure A3 presents estimates that explore potential mechanisms, finding no effects on other attitudes towards respondents' MCs, such as whether they are loyal or compromise too much. We also found no evidence that a positive experience in the trust game affected the amount of time respondents spent looking at their Members' votes.

party-consistent issue preferences than those assigned to the negative trust game experience (high affective polarization). This would be expected regardless of whether respondents saw information about how a copartisan or outpartisan MC voted, as partisans could also use an outpartisan MC's vote as a negative cue. To code the dependent variable, we compute the share of issues on which the respondent gave the same position as we had previously said their MC had taken; or, in the case when respondents had outpartisan MCs, the share of issues on which they gave the opposite position.

Figure 6 shows the results. We find no substantively or statistically significant effect of exogenously reducing affective polarization on whether respondents subsequently express party-consistent issue preferences in either Study 1 or Study 3. (This comparison corresponds with comparing issue preferences for those in Cells 2, 4, and 6 versus Cells 1, 3, and 5 of Table 3.) The point estimates are essentially zero, and the confidence intervals are very small. We continue to find null effects when we separately examine respondents who were of the same party or different party as their Member of Congress.

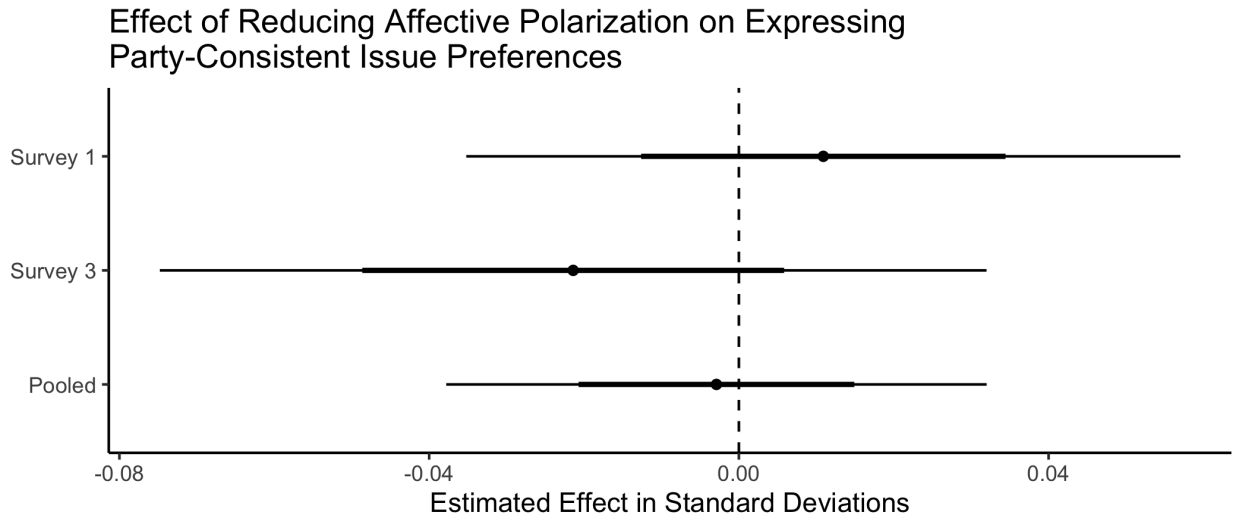
Support for Legislative Bipartisanship

Increases in affective polarization are also often blamed for partisan gridlock. For example, Levendusky (2018, p. 59) expresses concern that "affective polarization" may "make[] governance more difficult" by discouraging legislators from compromising.

To evaluate this prediction, Survey 2 deployed a manipulation developed by Harbridge and Malhotra (2011, Study 2). The authors found two Members of Congress, one of each party, who had each cast mostly party-line votes in one recent year but had cast a number of votes with the outparty in another recent year. The respondents are told about the Member of Congress of their party and it is randomly assigned whether they learn about the year of votes when this copartisan member cast party-line votes or often cast votes with the outparty.

After the vignette, respondents are asked whether they approve or disapprove of the job the representative is doing in Congress. Harbridge and Malhotra (2011) find that respondents on

Figure 6: Testing for Greater Receptivity to Party Cues: Effect of Inducing Affective Polarization on Party-Consistent Issue Preferences



Notes: Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Table A8.

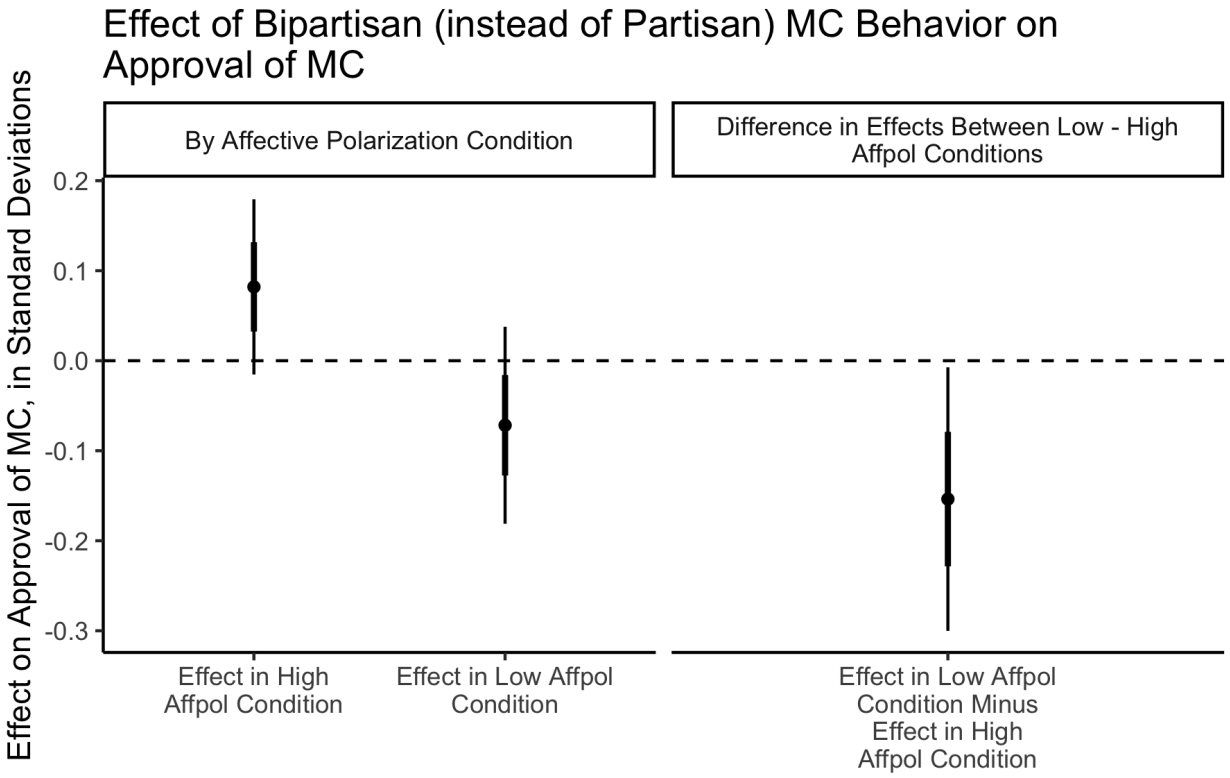
average approve of the job the representative is doing more when they learn about the year of votes where the member voted in a bipartisan manner.

We build on Harbridge and Malhotra (2011) by also randomly assigning respondents to either the positive or negative trust game experience before this vignette. If less affectively polarized voters would be more friendly towards bipartisanship, we should expect to see voters assigned to the reduced affective polarization condition additionally approve of the politician in the bipartisanship condition (or, equivalently, additionally disapprove of the politician in the partisan condition).

Figure 7 shows the results. The first coefficient shows the effect of the bipartisan condition for respondents in the high affective polarization condition. Similar to Harbridge and Malhotra (2011), we find that these respondents are $d = 0.31$ standard deviations more approving of the Member of Congress described as bipartisan instead of partisan. However, inconsistent with speculation in the literature on affective polarization, we do not find that respondents whose level

of affective polarization has been reduced desire more bipartisan behavior. If anything, as the second coefficient shows, respondents who are assigned to be more affectively polarized reward bipartisanship *less*: they show an increase of only $d = 0.19$ in approving of the job this MC is doing in Congress. The difference in these effects is not significant ($d = -0.11$, $p = 0.17$), but if anything indicates that respondents assigned to have decreased affective polarization rewarded bipartisan behavior *less*.

Figure 7: Testing for Decreased Support for Bipartisanship in Congress



Notes: See Figure A4 for condition means. Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Table A9.

Support for Democratic Norms

We next consider whether decreasing affective polarization bolsters support for democratic norms, as considerable research speculates. To test this, in Surveys 1, 2, and 4, after respondents played the trust game, we asked a wide variety of questions using multiple approaches to probe support for democratic norms. In particular, we asked respondents:

- An index probing opposition to democratic norms. This index included items such as "If a journalist accuses a [in party] politician of misconduct without naming their sources, the journalist should be criminally investigated." (See Appendix C.2.8 for the items.)

- Several vignettes from Lelkes and Westwood (2017). In each of these vignettes, respondents read a brief, fictional news article and were asked for their reaction:

  - *Suppression Vignette.* Respondents read a brief fictional news story with the headline "Police Use Tear Gas on Peaceful Young [Party] Protest." Respondents were asked their agreement with the decision to use tear gas.

  - *Corruption Tolerance Vignette.* The headline was "Donations from Millionaire Businessman to [Party] Super PACs in Question" and respondents were asked if they support an investigation.

  - *Antilocution Vignette.* The headline was "[Party]s Drive Congress to Do Less Than Last Year's Record-Breaking Low" and respondents were asked if they thought a news website should share this article.

  - *Election Override Vignette.* The headline was "Local [Party] Candidate Calls for State Legislature to Decide Election" and respondents were asked if they support the [Party] state legislature determining the outcome of the election.

  In Surveys 1 and 2, respondents were randomly assigned to only one of these vignettes. In Survey 4, respondents saw both the corruption tolerance and election override vignettes. In

some surveys, we randomized whether the respondent read about the in- or out-party while in other surveys this was fixed. Table B1 provides information for each survey whether the vignette party was in-party only, out-party only, or randomized. See Appendix C.2 for the vignettes' and questions' full text.

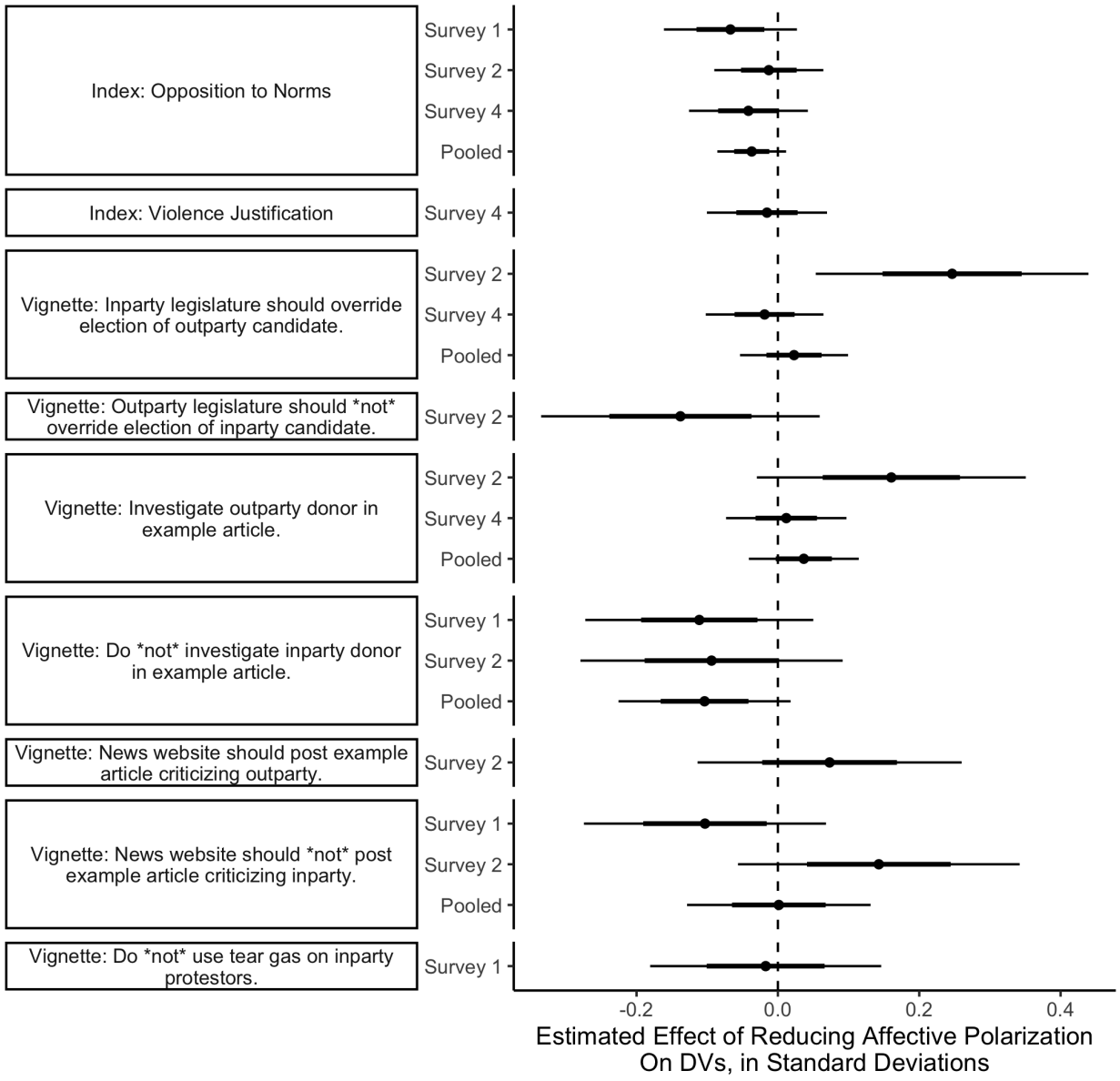
- A partisan violence justification index, four items from Kalmoe and Mason (2019). Statements include "How much would it be justified for [in party]s to use violence to advance their political goals these days?"

We estimate the effect of decreasing affective polarization on these norms indices and vignettes by regressing each outcome on an indicator for the randomly assigned positive trust game experience and pre-registered pre-treatment covariates (7-point partisanship scale and partisan strength). We code the direction of all the items such that higher values correspond with greater opposition to norms, harsher actions towards outpartisans, and leniency towards copartisans. Therefore, the literature's expectation would be for negative coefficients (i.e., that reducing affective polarization would make people less likely to oppose democratic norms).

In Figure 8, we find consistently small and statistically insignificant effects on each of the norms outcomes. We find no evidence that an exogenous decrease in affective polarization causes a downstream decrease in opposition to democratic norms. Even those estimates with the expected negative point estimates have tight confidence intervals that overlap zero. Appendix Figure A5 shows that the same pattern holds on the individual items on the norms index.[15] We likewise see no evidence of a diminished taste for norm violations in the vignettes; for example, participants in the decreased affective polarization condition are no less likely, and if anything more likely, to say they support an inparty-controlled legislature overturning the election of an outparty candidate. Although not all the point estimates are exactly zero, the spread of the point estimates around zero is consistent with what we would expect from sampling variability.

[15]Following Kingzette et al. (2021), we examined the effects on the norms scale by party and find nulls for both parties.

Figure 8: Testing for Undermined Support for Democratic Norms



Notes: Individual items in the norms index reported in Figure A5. Individual items in the violence index reported in Figure A6. Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Table A10.

Perceptions of Objective Conditions

The last set of outcomes we examine is whether a decrease in affective polarization leads to more accurate perceptions of objective conditions. Prior work has suggested that affective polarization strengthens the "perceptual screen" (Campbell et al. 1960) through which partisans are thought to see the world, leading Democrats and Republicans to perceive objective conditions differently. For example, Iyengar, Sood and Lelkes (2012, p. 428) hypothesize a connection between affective polarization and "partisan bias in perceptions of economic conditions."

We test this prediction in Survey 4 by asking respondents questions about their perceptions of unemployment rates and deaths from COVID-19. After respondents completed the trust game, we reminded respondents that "Donald Trump has been President of the United States for four years," then asked them what they thought the unemployment rate was and how many Americans had died of COVID-19 (see Appendix C.2 for wording). If affective polarization cause differences in stated perceptions of objective conditions, we should observe differences in how respondents answer these questions depending on whether they were randomly assigned to the positive or the negative trust game experience condition.

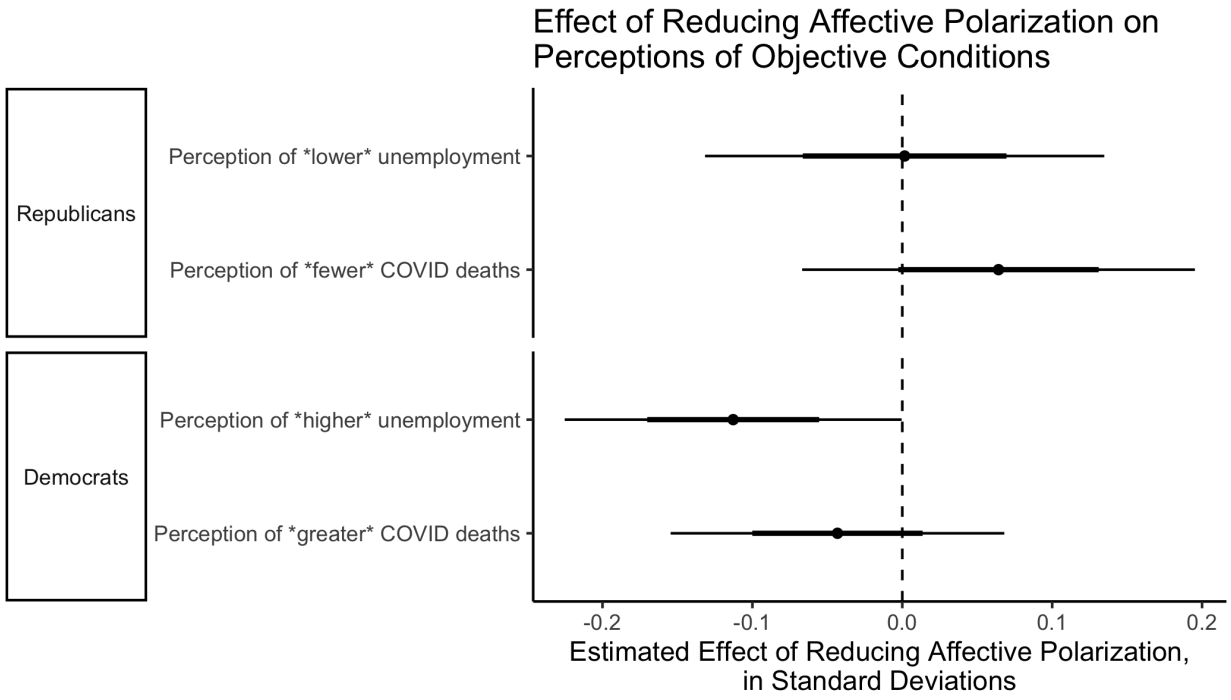
As expected, there were large partisan differences in stated perceptions of the number of COVID deaths and of the unemployment rate, with Republicans stating perceptions far lower, and Democrats stating perceptions far higher, of both. However, decreasing affective polarization did not appear to mitigate this partisan bias. Figure 9 shows these results.[16]

Alternative Explanations and Robustness Checks: Survey 5, Alternative Manipulations of Affective Polarization

Some may wonder whether the downstream effects of the changes in measured affective polarization the trust game produced are informative about the downstream effects of other

[16]The result that Democrats in the high affective polarization condition say they perceive higher levels of unemployment just meets the conventional threshold for statistical significance ($p = 0.03$) and does not survive adjusting for multiple comparisons.

Figure 9: Testing for Increased Bias in Perceptions of Objective Conditions



Notes: Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Table A11.

interventions that would change affective polarization. Might reducing affective polarization in other ways have salutary downstream effects even though we did not find them with this manipulation? This concern can be expressed more precisely in terms of four more specific concerns, each of which we help address with a final pre-registered survey, Survey 5.

The first concern is that the trust game and other interventions might manipulate affective polarization *among different populations* (i.e., different compliers). Our evidence so far does not support this concern: Table A1 shows that the trust game appears to have successfully manipulated affective polarization broadly, at every level of the 7-point partisan identification scale. However, Survey 5 addresses this concern further.

The second concern regards the exclusion restriction: perhaps a positive experience in the

trust game also affected constructs other than affective polarization and the effects of these other constructs cancelled out the salutary effects of reducing affective polarization. For this to have occurred, these other constructs would need to have *undermined* norms and accountability to an extent that exactly offset the salutary effects of reducing affective polarization. This seems doubtful; however, Survey 5 again addresses this concern further.

Third, one could be concerned that the trust game affected *measured* affective polarization but not *actual* affective polarization—i.e., perhaps the trust game induced non-random measurement error that led respondents to *answer the survey questions* differently even though *actual affective polarization* did not change. We *did* find downstream consequences of affective polarization for general social distance items (see Figure 3), and Westwood and Peterson (2020) also found downstream predicted effects of this manipulation on racial affect, suggesting this is unlikely. However, Survey 5 again addresses this further.

Fourth, an obtrusive manipulation such as the trust game may raise respondent suspicion and produce demand. For this to have led to null results that concealed the literature's predicted effects, a positive trust game experience would need to have led respondents to misrepresent their attitudes in a manner consistent with less accountability and being more anti-democratic, cancelling out the salutary effects of the reduced affective polarization. This seems unlikely as prior work shows that respondents are unlikely to alter their behavior on survey experiments to generate demand effects (Mummolo and Peterson 2019). Or, alternatively, it could be that the trust game did not actually manipulate affective polarization at all, but that respondents thought researchers wanted to see this occur. Survey 5 again addresses this further.

Survey 5 helps further address these concerns by examining the downstream effects of three additional approaches for manipulating affective polarization.[17] Ahler and Sood's (2018) intervention reduces affective polarization by correcting misperceptions about the demographic

[17]This was inspired by reviewer comments and by a working paper that appeared subsequent to our initial draft of this paper, Voelkel et al. (2021), which drew our attention to Lees and Cikara's (2020) manipulation, which we use in Survey 5.

composition of outpartisans. Lees and Cikara's (2020) intervention alters perceptions of outpartisans' political views on a policy issue. Finally, Levendusky's (2020) intervention prompts individuals to ruminate about an outpartisan friend. Survey 5 implements all three of these interventions. These interventions complement the trust game by manipulating affective polarization in three conceptually distinct manners.

Due to space constraints, we report Survey 5 in full (including the text of the interventions) in Appendix E and only discuss the main points here.

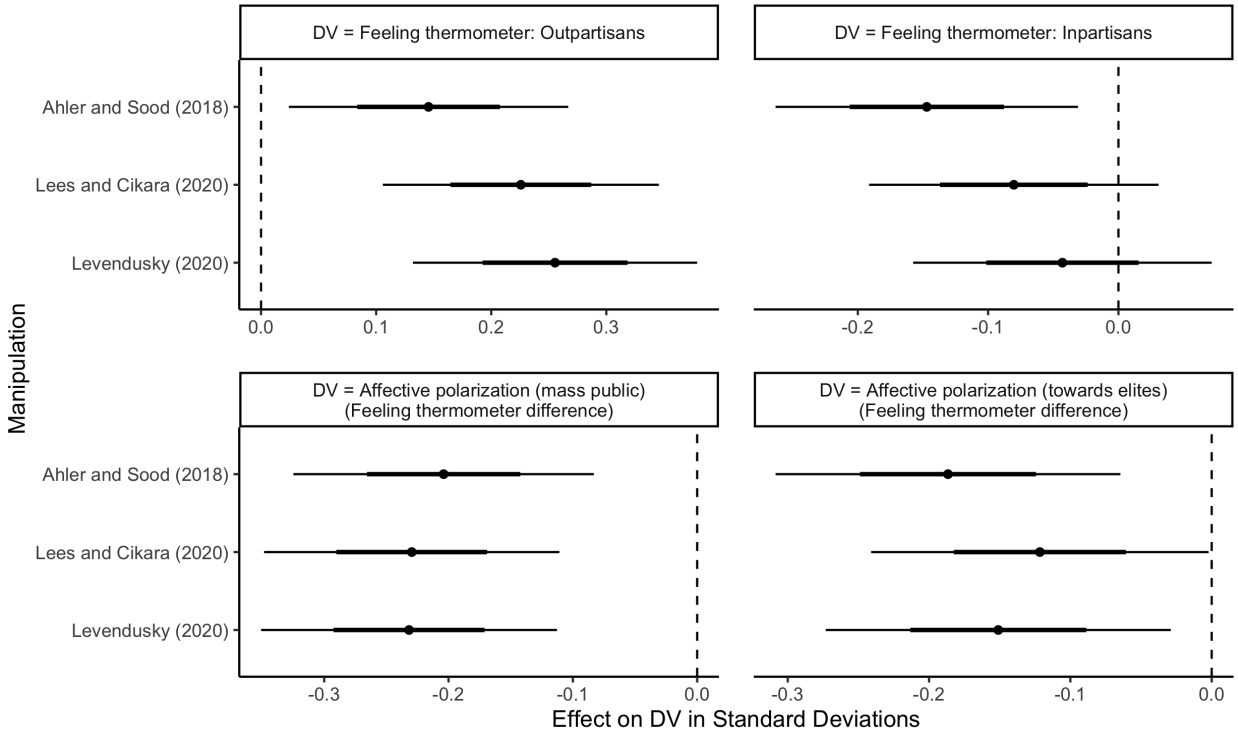
In June 2021, $N = 2,504$ respondents to the online survey vendor Dynata were randomly assigned to control, one of the three manipulations above, or the negative trust game manipulation. Survey respondents needed to meet the same eligibility criterias as Surveys 1-4.

Replicating the results reported by Ahler and Sood (2018), Lees and Cikara (2020), and Levendusky (2020), we report in Figure 10 that all three of these interventions reduced affective polarization by a statistically significant and similar degree compared to control. The interventions reduce affective polarization by 0.22 standard deviations on average ($t = -4.54$, $p < 0.001$), or 7.4 degrees. This is about half the size of the effect of the trust game manipulation (14.3 degrees), but still meaningful.

Given the smaller "first stage" effect of these manipulations on affective polarization, Survey 5 focused on downstream outcomes in the norms and social distance domains; these only required examining differences of means (whereas the other outcome areas rely on interaction terms, which are estimated with less statistical precision).

Figure 11 reports the results of Survey 5 on downstream political outcomes, estimating the effect of receiving any of the three new interventions versus the control group. We see consistent null results on political outcomes, as in Surveys 1-4. A seemingly unrelated regression model examining the joint significance of the effects of the treatments on the individual items in both domains likewise has a non-significant p-value of 0.65. Appendix Figure B5 shows that the results are similar for all three individual interventions.

Figure 10: Survey 5: Manipulation Checks



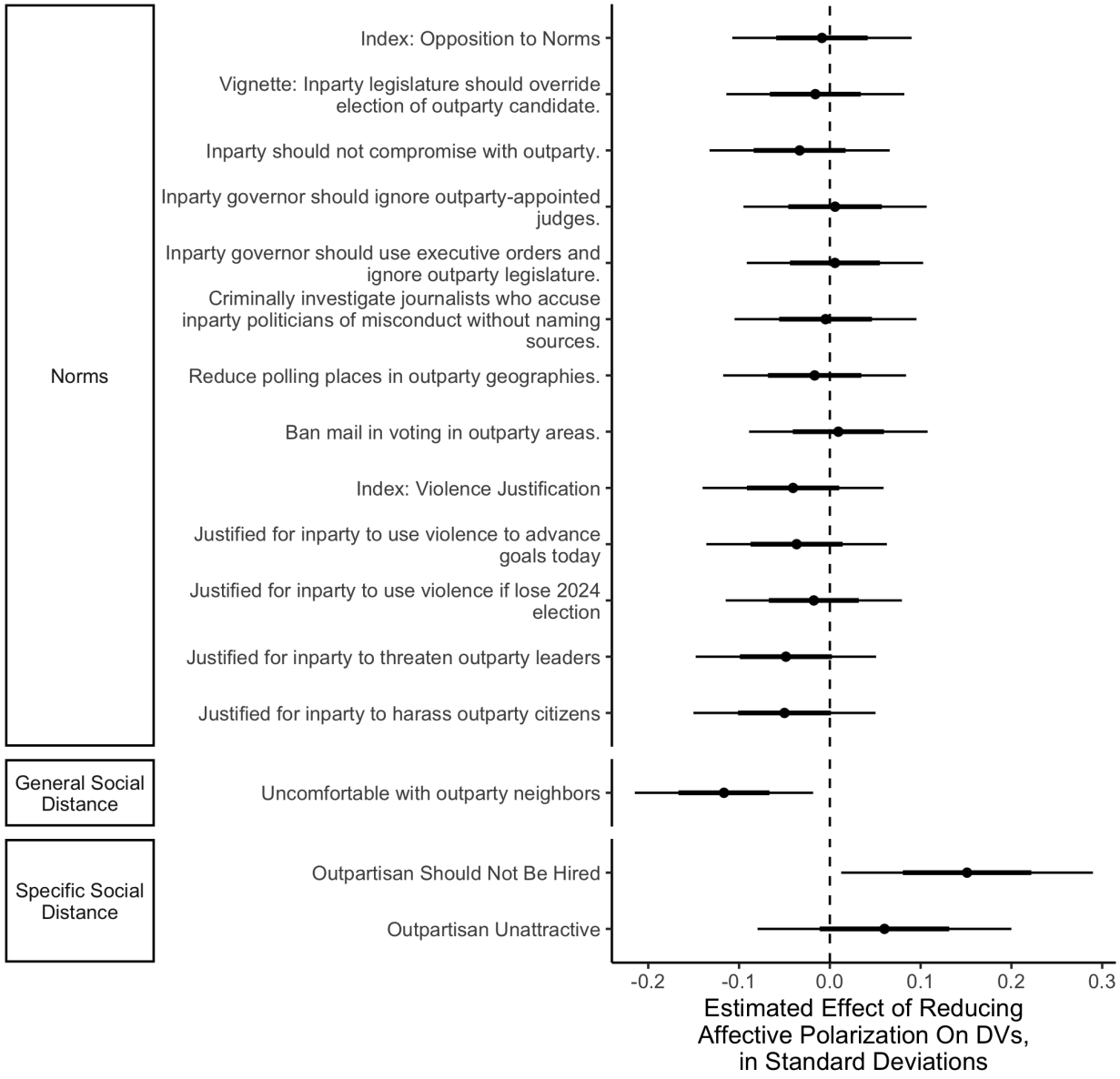
Notes: Point estimates are from multivariate regressions controlling for pre-registered covariates. Numerical values are presented in Appendix Table A12.

These results can help assuage concerns that our findings were particular to the trust game manipulation. Three additional, conceptually distinct approaches for reducing affective polarization produced similar results, and are unlikely to all have unusual compliers, violate the exclusion restriction in the same manner, or produce the same pattern of non-random measurement error.

As elaborated in Appendix E.2, Survey 5 also helps address the potential for demand effects by using interventions that are less overt, using additional distractor questions, and finding no signs of demand on open ended questions probing respondent's perceptions of the study's purpose.

Finally, consistent with Survey 5's evidence, recent papers by Voelkel et al. (2021), Santoro and Broockman (2021), and Levy (2021) also present interventions that reduce affective polarization but find no downstream impacts on a variety of political outcomes. Together our evidence and

Figure 11: Survey 5 Results: Average Effects of Alternative Approaches for Manipulating Affective Polarization



Notes: Point estimates are from multivariate regressions controlling for pre-registered covariates. Figure B5 shows that the results are similar for all three individual interventions. Numerical values are presented in Appendix Table A13.

these papers present a total of seven interventions to reduce affective polarization, none of which have downstream consequences for a wide variety of political outcomes. This bolsters confidence that this paper's results are not a false negative or an artifact of the questions we asked.

By contrast, we again see that the interventions reduced discomfort with having outpartisan neighbors, consistent with a decrease in general social distance. However, as discussed further in Appendix E, we also examined effects on two new more specific social distance items, which asked individuals to evaluate either the attractiveness of an outpartisan or whether they would hire an outpartisan worker. We do *not* find the anticipated decreases for these specific social distance items. This suggests that the causal effects of affective polarization even in the social distance area may not be robust to more specific questions and may only appear for more general survey questions. We return to this issue in the discussion section.

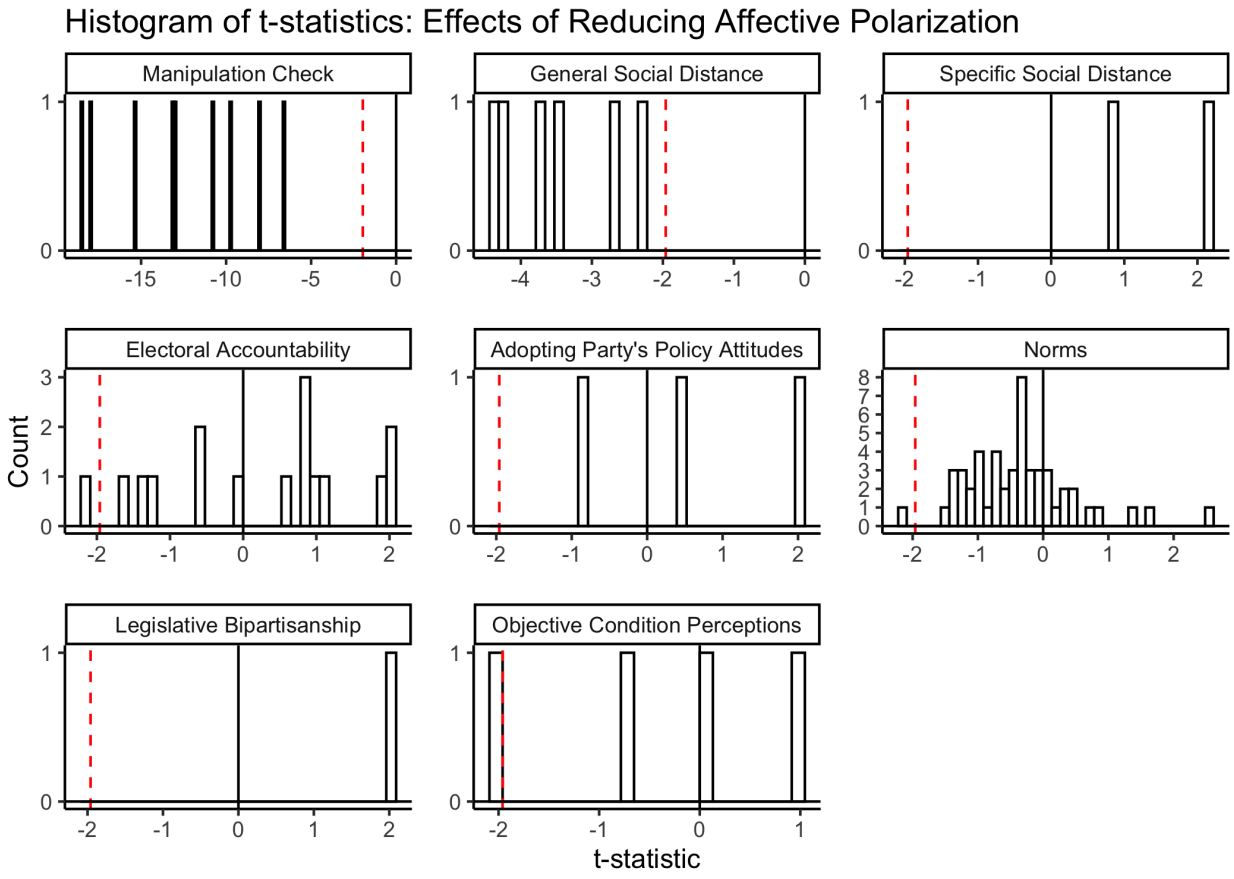
Summarizing the Results

We have presented many estimates of the downstream effects of reducing affective polarization across many outcomes. Contrary to the expectation of a great deal of literature, nearly all of our findings were null. To summarize our results and illustrate the consistency of our null findings, we computed the t-statistic associated with every result about the downstream effects of reducing affective polarization shown or discussed in the paper.

Figure 12 shows histograms of all these t-statistics, grouped by area. Each histogram also has a dashed red line at -1.96, the conventional statistical significance threshold (placed at -1.96 because all the coefficients are oriented such that the literature's speculation corresponds with negative coefficients).

The first two panels show that the manipulation check items and the general social distance items all have highly statistically significant results. The third shows the null results in the specific social distance area. Finally, the distribution of t-statistics in the five political outcome areas are consistent with null results. Out of the 75 t-statistics across these five areas, only 3 are large enough to just surpass statistical significance at the conventional level, in line with what would be expected by chance. Moreover each of these 3 statistically significant results are also accompanied by other clearly null results on either replications or closely related outcomes or tests, consistent

Figure 12: Histogram of t-statistics for results discussed and presented in paper, by outcome area



Notes: Histograms of t-statistics for results presented or discussed in the paper, organized by outcome area. The red vertical line is at -1.96, the conventional statistical significance threshold.

with them being false positives.[18] Finally, when applying a False Discovery Rate adjustment to all the p-values associated with all the t-statistics, all the manipulation check and general social distance items remain significant at the conventional threshold, but none of the formerly significant outcomes in the other areas do.

These consistently null results cannot be explained by limited statistical power. Not only was

[18]The first significant result is on the Survey 1 "investigate journalists" norms item; however, this did not replicate in Survey 2 or 4 (see Figure A5). The second significant result is the decreased reward Members of Congress earned for casting congruent votes in Survey 3 (see Figure 5c), but the results for incongruent votes in both Survey 1 and Survey 3 (see Figures 5a and 5b) go in the opposite direction. The third significant result is the increased perception of the unemployment rate among Democrats, but we do not see these results for their perceptions of COVID deaths, nor for Republicans on either outcome (see Figure 9).

our manipulation of affective polarization powerful (equal to approximately three decades of increased affective polarization) and our sample sizes large, we would expect to observe a larger proportion of statistically significant t-statistics were the null hypothesis wrong.

Comparing Experimental and Endogenous Estimates

We next show that the null results our experimental approach surfaced stand in contrast to large and positive—but misleading—results one would reach without it. In particular, for every result we discussed or presented in the paper, we computed (1) the experimental estimate we show in the paper (on standardized versions of each outcome) to (2) a non-causal endogenous estimate, computed by estimating the observational relationship between the difference in affective polarization equal to the size our experimental treatment created and each standardized outcome. This allows us to compare endogenous estimates of the form that the existing literature would compute to our causal estimates on the same scale.

Figure 13 compares these endogenous estimates to our experimental estimates. The first panel, 13a, shows the general social distance outcomes. In this domain, where our argument predicts and we find causal effects of affective polarization, the endogenous correlational estimates and the experimental estimates are nearly identical.

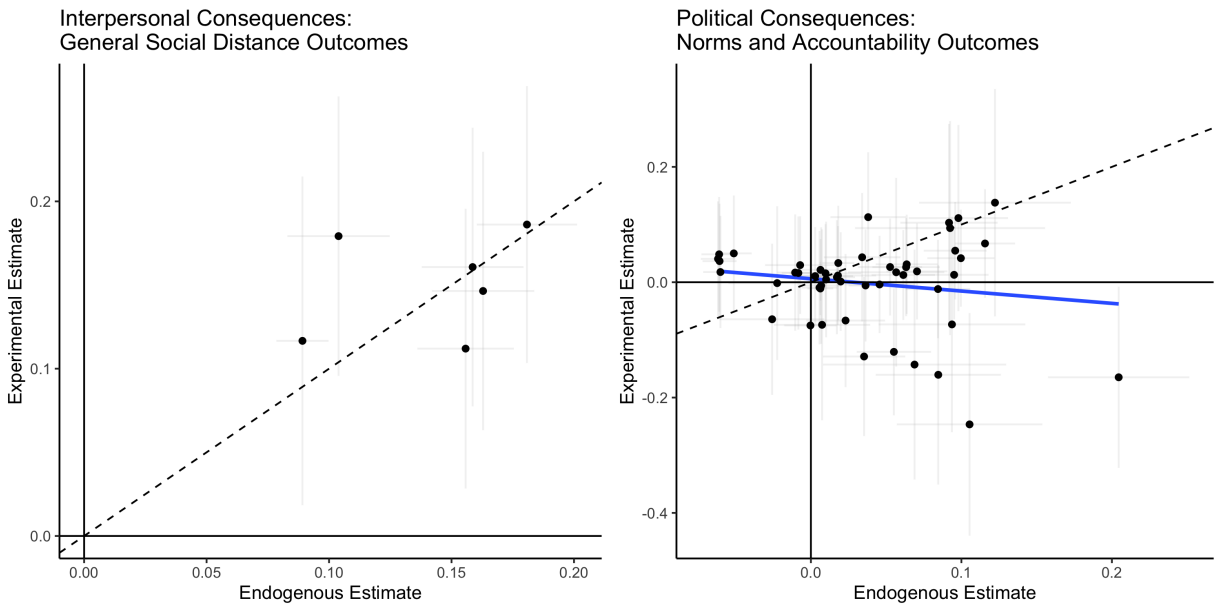
Figure 13b shows the same relationship for the political outcomes. Approximately 79% of the endogenous estimates are positive, meaning that a version of this paper which followed existing literature in examining the observational relationship between affective polarization and these political outcomes would have reported nearly all positive results. However, only 46% of the experimental estimates are positive, essentially a coin flip. Moreover, 63% of the endogenous estimates are larger than the experimental estimates: the endogenous estimates consistently overestimate the causal effects of affective polarization. Finally, the blue line of best fit shows that there is no relationship between the size of the experimental and the endogenous estimates.

These results provide evidence for the alternative relationship between affective polarization and downstream political choices we hypothesized in Figure 1b. Affective polarization correlates

Figure 13: Comparing Endogenous Estimates (x-axes) with Experimental Estimates (y-axes)

(a) Social Distance Outcomes

(b) Political Outcomes



Notes: The x-axis value corresponds with the experimental estimate presented in the paper on standardized versions of each outcome. The y-axis value is a non-causal endogenous estimate of the same effect, computed by estimating the observational relationship between an "increase" in affective polarization equal to the size of the effect of the trust game and each standardized outcome. 95% confidence intervals in both regressions are shown surrounding the point estimates. The dotted line is a 45 degree line showing the slope $y = x$. Panel 13b shows a line of best fit in blue.

with a number of other constructs, including the intensity of partisanship, policy preferences, and many more, perhaps as it may summarize these considerations. However, because these other constructs are also likely to correlate with downstream political choices, they represent omitted variables that inflate the correlation between affective polarization and these downstream choices. These results provide caution for analysts interested in affective polarization's impacts, as it makes it difficult, if not impossible, for analysts to isolate the downstream causal effects of affective polarization from the effects of these other considerations in observational data, and very likely to overestimate them.

Discussion

Across many democratic societies, citizens express increasing dislike for those of opposing political parties relative to their own (Gidron, Adams and Horne 2020). Scholars routinely speculate that this affective polarization has adverse downstream consequences for interpersonal and political decisions (see Table 1 for review). However, little theory or data has rigorously considered these concerns.

Returning to foundational theories of the role of affect in judgment, we argued that we should not expect affective polarization to influence all judgments somewhat related to party, but principally those judgments where individuals do not have other, more relevant considerations they would more readily call to mind. Moreover, we pointed out that, for affective polarization to alter judgments, it creates a trade-off: individuals must give up whatever it is they preferred about the choice that affective polarization made them abandon.

For general survey questions about interpersonal topics, such as whether someone would like having an outpartisan neighbor, we argued that affective polarization may be more likely to have an impact: there are no trade-offs, and few if any other dimensions are available for respondents to rely on. Our results were consistent with our argument: when we leveraged multiple approaches for introducing exogenous variation in affective polarization, they all produced downstream changes in response to general social distance survey questions.

However, when it comes to political judgments, our argument suggested greater skepticism—and we found no evidence that exogenously manipulating affective polarization had downstream consequences. We found these nulls across five issues areas where we leveraged a number of measurement approaches, including matching respondents to their actual representatives and examining both levels of party loyalty and willingness to hold them accountable for their votes.

Limitations and Potential Alternative Explanations

Our research is the first to rigorously examine the downstream consequences of affective polarization, but it should not be the last; our research has several important limitations.

One important limitation of our research is that we only examine the immediate, short-run effects of affective polarization on outcomes that can be measured in a survey. It is possible that there are other long-run or indirect effects of affective polarization that our research design cannot measure. For example, if affective polarization reduces interactions with outpartisans and conversations with outpartisans expose people to rationales for the other party's policies, affective polarization could have indirect effects on people's political attitudes even if it does not have direct effects on them. This is worthy of future study.

In addition, while we tested for downstream effects on the outcomes most commonly named in the literature, future research can use the paradigms we present to test for effects on additional outcomes. Likewise, studies that manipulate affective polarization in the field or measure behavioral outcomes would also be welcome. We also did not explicitly test for the mechanisms that explain why respondents might not rely on affect in decision making. This remains an important next step.

A third limitation is that the manipulations primarily reduced outpartisan dislike rather than inpartisan support. This is consistent with what has driven the increase in affective polarization over time (i.e., increasing dislike for the outparty Iyengar, Sood and Lelkes 2012). Manipulating inpartisan support is therefore another possible avenue for future research. Future research could also examine the downstream effects of *further increasing* affective polarization, although this may raise ethical concerns.

Implications for Future Research

Our results suggest several implications for future research and concrete implications for those seeking to improve American democracy.

First, our work suggests that future research on the political implications of affective polarization may wish to further examine the robustness of the downstream effects of affective polarization in the interpersonal domain. Although we found evidence of downstream effects on general interpersonal items, we found in Survey 5 that these effects did not manifest when we asked respondents to make judgments about specific individuals. (This is not to say that we found no partisan bias, but rather that affective polarization did not influence the magnitude of this bias, as is often assumed.) It is possible that this finding speaks to a broader distinction between affective polarization's downstream effects for specific judgments versus general survey questions—perhaps because specific judgments involve trade-offs that answers to general survey questions usually do not. Our work is too preliminary to form firm conclusions on this point, and we plan to take this question up in future research focusing on more general political attitudes (versus the largely specific political choices we often posed to respondents in our surveys).

Second, insofar as we find some evidence of downstream effects in the social domain, future research may wish to more closely examine those areas where the interpersonal domain and the political domain are inherently linked. This could be a fruitful and important area for study that holds important normative implications. For example, we found that reducing affective polarization made individuals more willing to discuss politics with outpartisan citizens—deliberation which, if it were to occur, could have important consequences our work was not able to measure.

Finally, our results have implications for efforts to improve democracy. There are a number of hypotheses for what is behind negative trends in democratic societies, such as the decline in ticket splitting in the United States (e.g., Hopkins 2018) or the decay of certain democratic norms (e..g, Graham 2021). In order to reverse these trends, researchers must build an understanding of which factors are contributing to these trends and which are not. We see our research as a part of this effort. It suggests that attempting to reduce affective polarization—although potentially valuable for many reasons—may not be the most effective way to reverse these trends. Our findings in

Survey 5 that three common interventions that reduce affective polarization do not increase support for democratic norms underscore this implication. Our findings in no way end the conversation about the potential political impacts of affective polarization on the functioning of American democracy, but rather represent a step towards more focused theorizing and rigorous empirical study of this important topic. Moreover, our findings in no way discount the role of *other* features of partisanship, such as partisan identity, in potentially contributing to negative trends in democracy. For example, Klar (2013) and Graham (2021) both find evidence that partisan *identity* (which is distinct from affect) has downstream consequences for political attitudes, including willingness to punish politicians who violate democratic norms—unlike interventions to reduce affective polarization. Indeed, in casting doubt on the role of affective polarization, our work underscores the need to attend to these other possibilities. We hope that our work stimulates additional rigorous attention on what will—and will not—bolster American democracy.

References

Abramowitz, Alan I and Steven Webster. 2016. "The rise of negative partisanship and the nationalization of US elections in the 21st century." *Electoral Studies* 41:12–22.

Ahler, Douglas J. and Gaurav Sood. 2018. "The parties in our heads: Misperceptions about party composition and their consequences." *Journal of Politics* 80(3):964–981.

Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Berg, Joyce, John Dickhaut and Kevin McCabe. 1995. "Trust, reciprocity, and social history." *Games and economic behavior* 10(1):122–142.

Brody, Richard A. and Paul M. Sniderman. 1977. "From life space to polling place: The relevance of personal concerns for voting behavior." *British Journal of Political Science* 7(3):337–360.

Campbell, Angus, Philip E. Converse, Warren E. Miller and Donald E. Stokes. 1960. *The American Voter*. New York: John Wiley.

Druckman, James N. and Matthew S. Levendusky. 2019. "What do we measure when we measure affective polarization?" *Public Opinion Quarterly* 83(1):114–122.

Druckman, James N, Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2020*a*. "Affective polarization, local contexts and public opinion in America." *Nature Human Behaviour* pp. 1–11.

Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky and John Barry Ryan. 2020*b*. "How Affective Polarization Shapes Americans' Political Beliefs: A Study of Response to the COVID-19 Pandemic." *Journal of Experimental Political Science* pp. 1–12.

Finkel, Eli J., Christopher A. Bail, Mina Cikara, Peter H. Ditto, Shanto Iyengar, Samara Klar, Lilliana Mason, Mary C. McGrath, Brendan Nyhan, David G. Rand et al. 2020. "Political sectarianism in America." *Science* 370(6516):533–536.

Freeder, Sean. 2020. "It's No Longer the Economy, Stupid: Selective Perception and Attribution of Economic Outcomes." Available at `https://seanfreeder.files.wordpress.com/2020/12/paper_economy.pdf`.

Gidron, Noam, James Adams and Will Horne. 2020. *American Affective Polarization in Comparative Perspective*. Cambridge.

Graham, Matthew H. 2021. "Does Partisan Identity Reduce Support for Electoral Fairness?" Available at `https://osf.io/vwe36/`.

Graham, Matthew H and Milan W Svolik. 2020. "Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States." *American Political Science Review* 114(2):392–409.

Harbridge, Laurel and Neil Malhotra. 2011. "Electoral incentives and partisan conflict in Congress: Evidence from survey experiments." *American Journal of Political Science* 55(3):494–510.

Hopkins, Daniel J. 2018. *The increasingly United States: How and why American political behavior nationalized*. Chicago.

Iyengar, Shanto, Gaurav Sood and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76(3):405–431.

Iyengar, Shanto and Masha Krupenkin. 2018. "Partisanship as social identity; Implications for the study of party polarization." *The Forum* 16(1):23–45.

Iyengar, Shanto and Sean J. Westwood. 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59(3):690–707.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean J. Westwood. 2019. "The origins and consequences of affective polarization in the United States." *Annual Review of Political Science* 22:129–146.

Kahneman, Daniel and Shane Frederick. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and biases: The psychology of intuitive judgment*, ed. Thomas Gilovich, Dale Griffin and Daniel Kahneman. Cambridge pp. 49–81.

Kalmoe, Nathan P. and Lilliana Mason. 2019. "Lethal mass partisanship: Prevalence, correlates, and electoral contingencies." Available at `https://www.dannyhayes.org/uploads/6/9/8/5/69858539/kalmoe___mason_ncapsa_2019_-_lethal_partisanship_-_final_lmedit.pdf`.

Kingzette, Jon, James N. Druckman, Samara Klar, Yanna Krupnikov, Matthew Levendusky and John B. Ryan. 2021. "How Affective Polarization Undermines Support for Democratic Norms." *Public Opinion Quarterly* .

Klar, Samara. 2013. "The influence of competing identity primes on political preferences." *Journal of Politics* 75(4):1108–1124.

Klar, Samara and Yanna Krupnikov. 2016. *Independent politics*. Cambridge.

Klar, Samara, Yanna Krupnikov and John Barry Ryan. 2018. "Affective polarization or partisan disdain? Untangling a dislike for the opposing party from a dislike of partisanship." *Public Opinion Quarterly* 82(2):379–390.

Lees, Jeffrey and Mina Cikara. 2020. "Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts." *Nature Human Behaviour* 4(3):279–286.

Lelkes, Yphtach and Sean J. Westwood. 2017. "The limits of partisan prejudice." *Journal of Politics* 79(2):485–501.

Levendusky, Matthew S. 2018. "Americans, not partisans: Can priming American national identity reduce affective polarization?" *Journal of Politics* 80(1):59–70.

Levendusky, Matthew S. 2020. "Our common bonds: Using what Americans share to help bridge the partisan divide." *Unpublished Manuscript* .

Levy, Ro'ee. 2021. "Social media, news consumption, and polarization: Evidence from a field experiment." *American economic review* 111(3):831–70.

Little, Andrew T., Keith Schnakenberg and Ian R. Turner. 2020. "Motivated reasoning and democratic accountability." Working paper, available at `https://osf.io/preprints/socarxiv/esfy6/`.

Mason, Lilliana. 2014. ""I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization." *American Journal of Political Science* .

McCoy, Jennifer and Murat Somer. 2019. "Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies." *The Annals of the American Academy of Political and Social Science* 681(1):234–271.

Mummolo, Jonathan and Erik Peterson. 2019. "Demand effects in survey experiments: An empirical assessment." *American Political Science Review* 113(2):517–529.

Ottati, Victor C. and Robert S. Wyer. 1993. Affect and Political Judgment. In *Explorations in political psychology*, ed. Shanto Iyengar and William J. McGuire. Duke pp. 296–316.

Pierson, Paul and Eric Schickler. 2020. "Madison's Constitution Under Stress: A Developmental Analysis of Political Polarization." *Annual Review of Political Science* 23:37–58.

Santoro, Erik and David Broockman. 2021. "The short-term, circumscribed, and conditional effects of cross-partisan conversation." Available at `https://osf.io/km3q2`.

Slovic, Paul, Melissa L. Finucane, Ellen Peters and Donald G. MacGregor. 2002. The affect heuristic. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. Thomas Gilovich, Dale Griffin and Daniel Kahneman. Cambridge.

Voelkel, Jan G., James Chu, Michael N. Stagnaro, Joseph S. Mernyk, Sophia Pink, Chrystal Redekopp, James Druckman, David Rand and Robb Willer. 2021. "Interventions Reducing Affective Polarization Do Not Improve Anti-Democratic Attitudes." Available at `https://osf.io/preprints/7evmp/`.

Westwood, Sean J. and Erik Peterson. 2020. "The Inseparability of Race and Partisanship in the United States." *Political Behavior* pp. 1–23.

Online Appendix

Contents

A Ethical Statement **48**

B Additional Tables and Figures **49**

B.1 Quantile Regression Results 66

C Surveys **68**

C.1 Survey Contents and Order 68

C.2 Question Text 69

C.3 Text of Additional Manipulations (Survey 5) 94

C.4 Survey Demographics and Representativeness 98

C.5 Attrition 100

D Supplemental Data on Trust Game **102**

E Survey 5: Additional Details and Results **105**

E.1 Results 106

E.2 Testing for Demand Effects 109

F References for Online Appendix **111**

A Ethical Statement

With respect to ethics, all of our surveys were IRB approved and we obtained voluntary and informed consent from all subjects. All subjects were aware they were taking part in a research study. Our research design required the use of deception. Participants were aware that deception might be used and we debriefed all subjects immediately following the conclusion of the survey. The survey vendor offered a predetermined, fair level of compensation to participants based on the length of the survey.

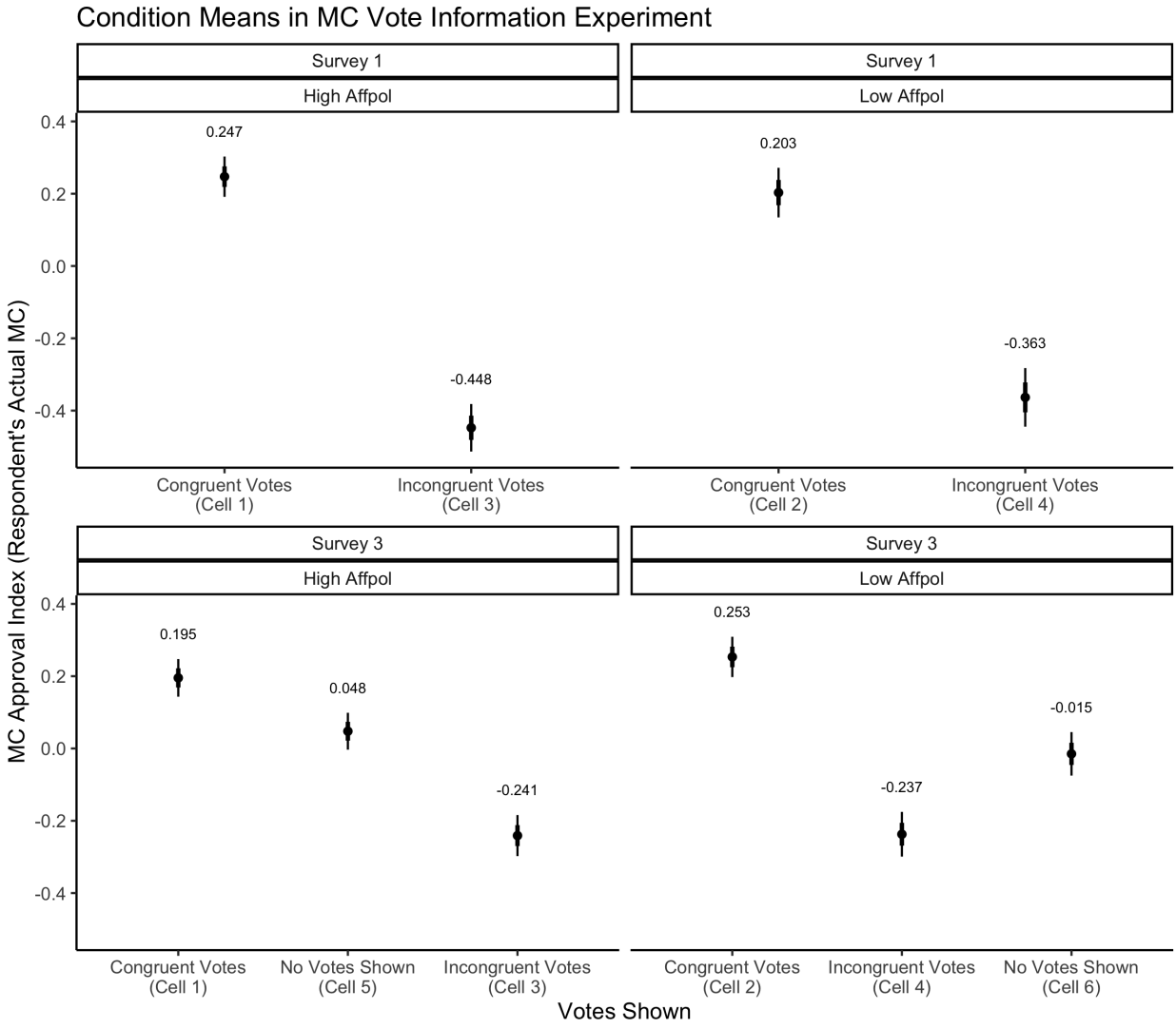
B Additional Tables and Figures

Table A1: Effect of Positive Trust Game Experience on Affective Polarization Towards Mass Public and Towards Elites, by Strength of Party ID

	DV = Outparty - Inparty Mass Public Feeling Therms	DV = Outparty - Inparty Elite Feeling Therms
PID = Not Strong Democrat	-22.2***	-22.0***
	(2.85)	(2.86)
PID = Lean Democrat	-27.8***	-24.4***
	(2.94)	(2.92)
PID = Lean Republican	-28.5***	-29.4***
	(3.42)	(3.51)
PID = Not Strong Republican	-29.1***	-30.8***
	(3.26)	(3.31)
PID = Strong Republican	-3.26	-1.45
	(3.13)	(3.07)
Strong Democrat × Positive Game Experience	-15.7***	-9.86***
	(2.65)	(2.88)
Not Strong Democrat × Positive Game Experience	-13.9***	-10.2**
	(3.07)	(3.24)
Lean Democrat × Positive Game Experience	-10.2**	-4.46
	(3.37)	(3.42)
Lean Republican × Positive Game Experience	-10.8*	-5.21
	(4.34)	(4.59)
Not Strong Republican × Positive Game Experience	-16.5***	-11.2**
	(3.44)	(3.71)
Strong Republican × Positive Game Experience	-18.0***	-12.6***
	(3.69)	(3.71)
Constant	52.7***	54.1***
	(1.76)	(1.87)
Observations	2135	2135

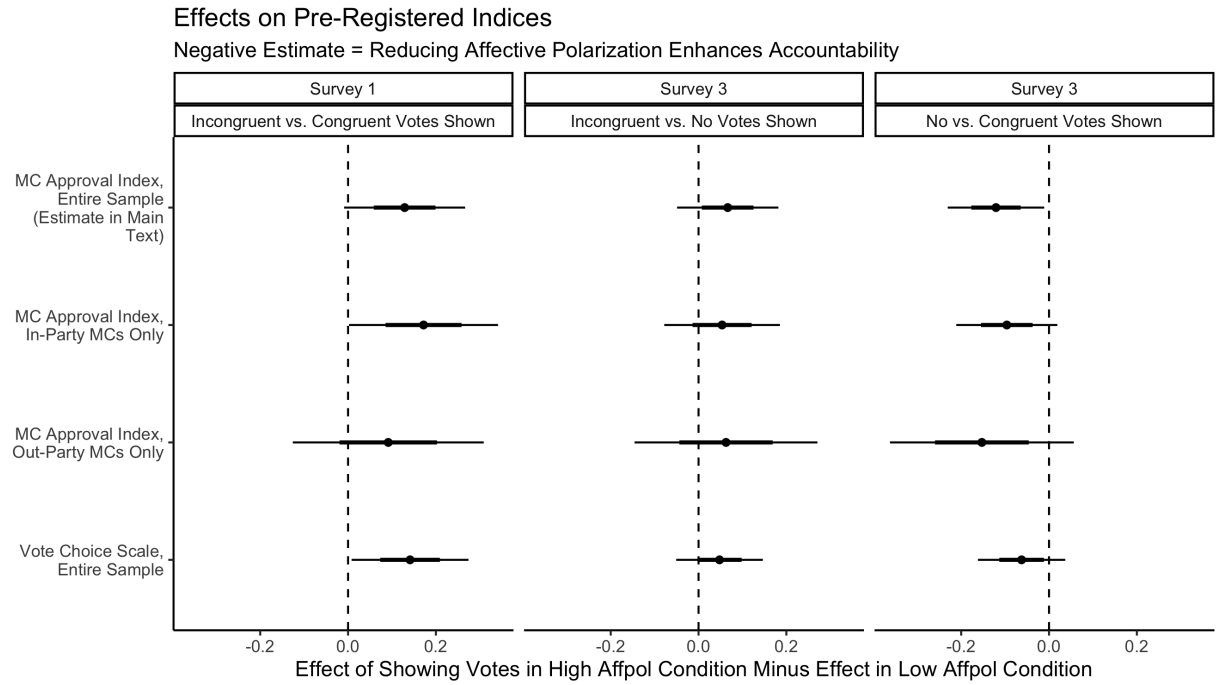
Notes: The rows with interactions show the effect of a positive game experience by party ID. The rows without interactions show differences between levels of affective polarization between different partisan groups within the control (negative game experience) condition, with PID = Strong Democrat as the omitted baseline category. Data are from Survey 4, as Survey 4 was the only survey that asked about affective polarization towards elites. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure A1: Condition Means in MC Vote Choice Experiment



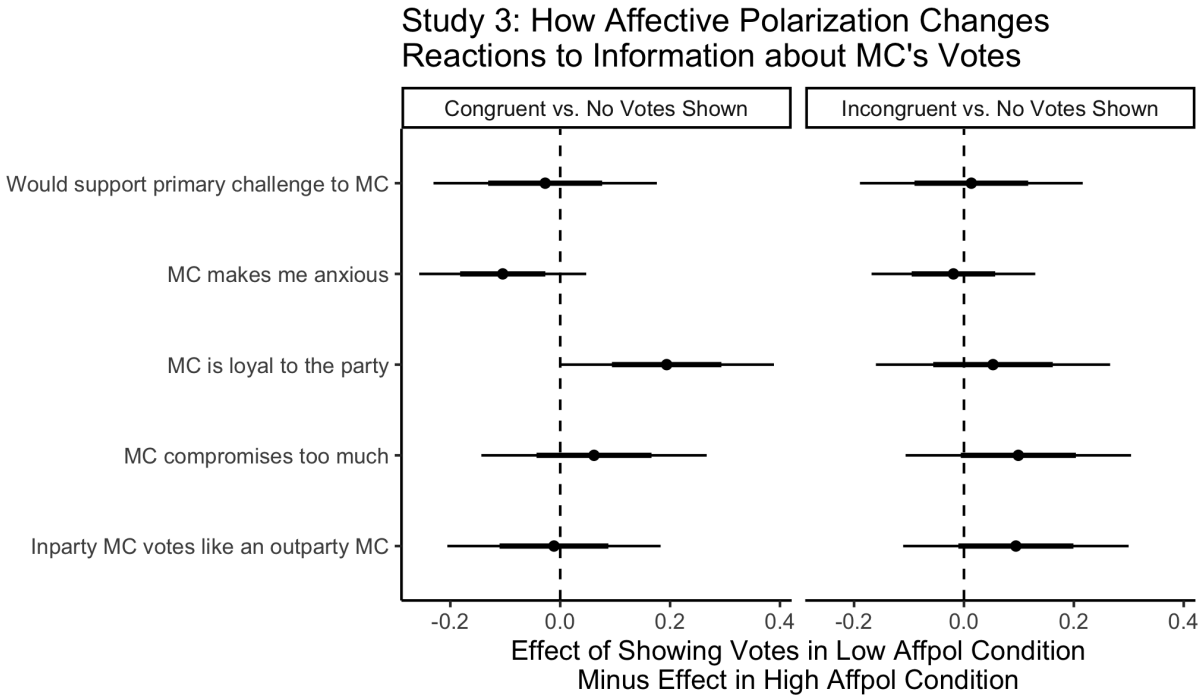
Notes: The Cell numbers in the coefficient labels refer to cells in Table 3. This Figure shows the condition means in the experiment reported in Figure 5 in the main text. Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are predicted probabilities from multivariate regressions controlling for pre-registered covariates.

Figure A2: Testing for Desensitization: Robustness



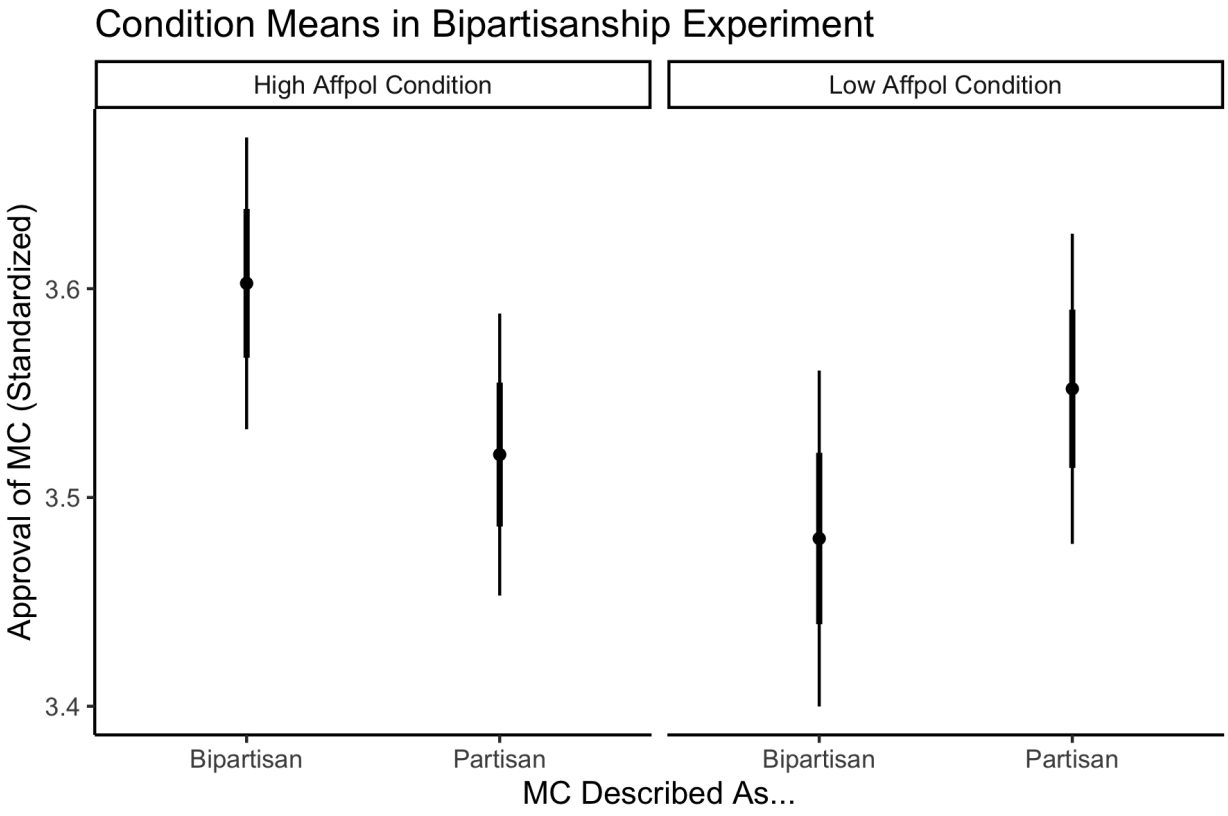
Notes: This Figure helps assess the robustness of the results reported in Figure 5 in the main text. The coefficients shown capture the interaction between the affective polarization and MC vote treatments; the first row in each panel reports the same coefficient as shown in the main text (i.e., the rightmost coefficients in each panel in Figure 5). The next two rows show the estimates for respondents with in-party MCs only and the following row the estimates for respondents with out-party MCs only, respectively. Under the hypothesis that reducing affective polarization makes individuals more willing to punish inparty MCs for incongruent votes, we should expect to see a negative coefficient in the middle "Incongruent vs. No Votes Shown" panel in the second row for "In-Party MCs Only." Under the hypothesis that reducing affective polarization makes individuals more willing to reward outparty MCs for congruent votes, we should expect to see a negative coefficient under the third panel, "No vs. Congruent Votes Shown" in the third row for "Out-Party MCs Only." The final row the results in the entire sample when using just the vote choice scale as an outcome instead of the entire MC approval index. Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates.

Figure A3: Testing for Desensitization: Potential Mechanisms



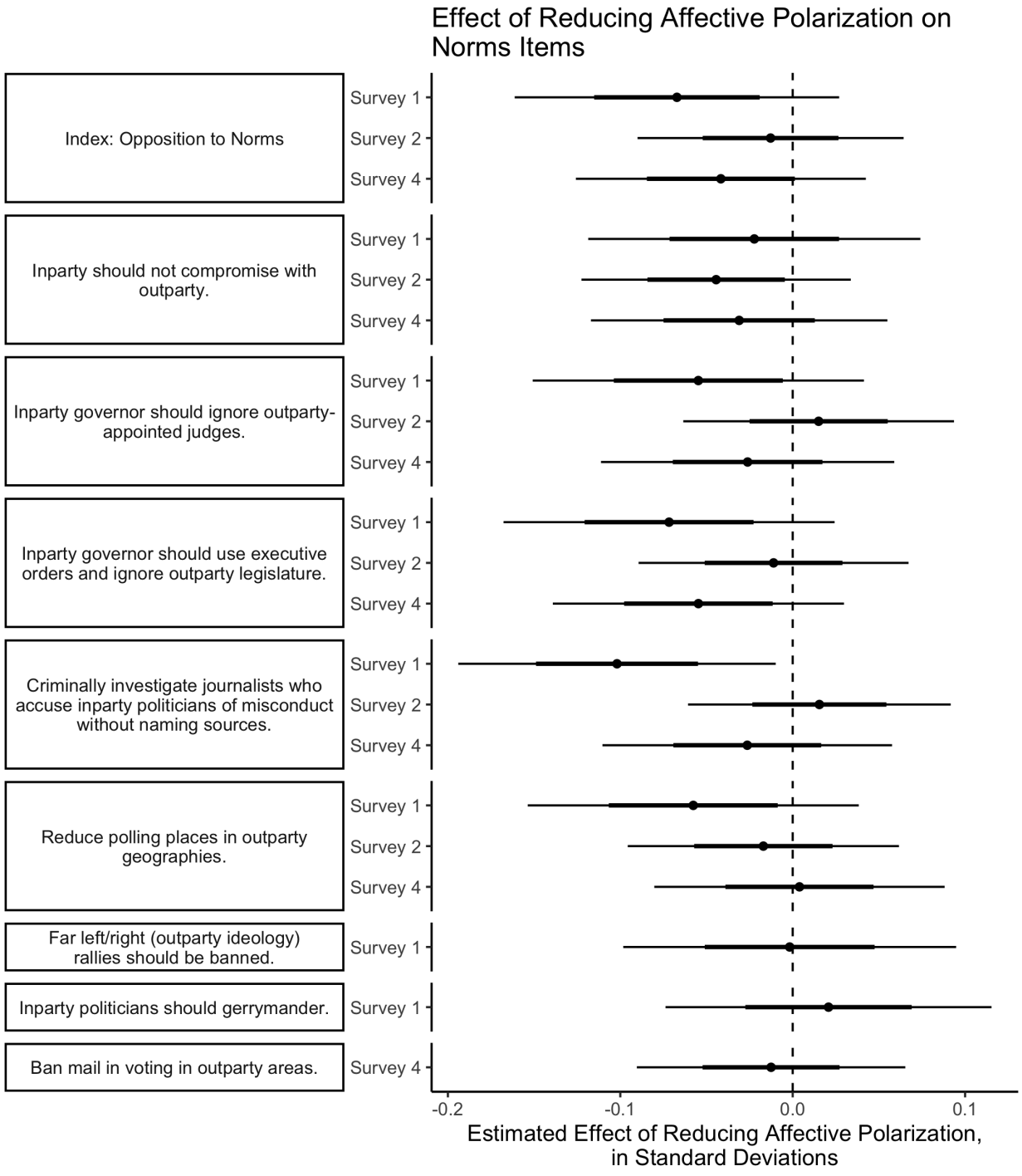
Notes: This Figure reports the difference-in-differences estimate from Figures 5b and 5c in the main text but on other "mechanism" items. Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates.

Figure A4: Condition Means in Bipartisanship Experiment



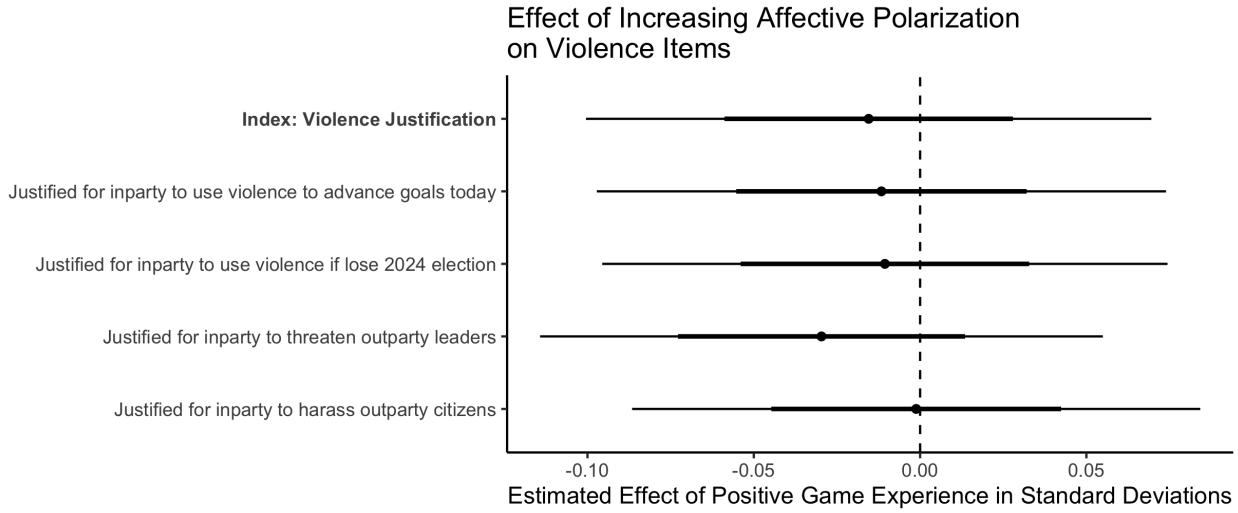
Notes: This Figure shows the condition means in the bipartisanship experiment reported in Figure 7 in the main text. Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates.

Figure A5: Effect of Increasing Affective Polarization on Individual Items in Norms Index



Notes: The first set of coefficients reproduces the estimates on the Norms index shown in Figure 8 in the main text. The remaining coefficients show estimates on the individual items in this index. For the full text of these items, see Appendix C.2.8. Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates.

Figure A6: Effect of Increasing Affective Polarization on Individual Items in Violence Justification Index



Notes: The first set of coefficients reproduces the estimates on the Violence Justification index shown in Figure 8 in the main text. The remaining coefficients show estimates on the individual items in this index. Point estimates are surrounded by one standard error (thick tails) and 95% confidence intervals (thin tails). Point estimates are from multivariate regressions controlling for pre-registered covariates.

Table A2: Effect of Positive Trust Game Experience on Manipulation Checks (Figure 2)

Outcome	Study	Effect	SE	CI
Feeling thermometer: Inpartisans	Survey 1	-0.040	0.046	-0.13, 0.05
Feeling thermometer: Inpartisans	Survey 2	0.077	0.038	0.003, 0.151
Feeling thermometer: Inpartisans	Survey 3	0.052	0.032	-0.011, 0.115
Feeling thermometer: Inpartisans	Survey 4	0.082	0.040	0.004, 0.16
Feeling thermometer: Outpartisans	Survey 1	0.457	0.047	0.365, 0.549
Feeling thermometer: Outpartisans	Survey 2	0.688	0.037	0.615, 0.761
Feeling thermometer: Outpartisans	Survey 3	0.575	0.032	0.512, 0.638
Feeling thermometer: Outpartisans	Survey 4	0.634	0.041	0.554, 0.714
Affective polarization (mass public) (Feeling thermometer difference)	Survey 1	-0.363	0.045	-0.451, -0.275
Affective polarization (mass public) (Feeling thermometer difference)	Survey 2	-0.485	0.037	-0.558, -0.412
Affective polarization (mass public) (Feeling thermometer difference)	Survey 3	-0.404	0.031	-0.465, -0.343
Affective polarization (mass public) (Feeling thermometer difference)	Survey 4	-0.428	0.040	-0.506, -0.35
Affective polarization (towards elites) (Feeling thermometer difference)	Survey 4	-0.272	0.041	-0.352, -0.192
Treated fairly in trust game (1-7 Likert Scale)	Survey 1	1.449	0.033	1.384, 1.514
Treated fairly in trust game (1-7 Likert Scale)	Survey 2	1.519	0.025	1.47, 1.568
Treated fairly in trust game (1-7 Likert Scale)	Survey 3	1.458	0.023	1.413, 1.503
Treated fairly in trust game (1-7 Likert Scale)	Survey 4	1.525	0.027	1.472, 1.578

Table A3: Effect of Increasing Affective Polarization on Social Distance Items (Figure 3)

Outcome	Study	Effect	SE	CI
Uncomfortable talking politics with outparty	Survey 4	-0.179	0.043	-0.263, -0.095
Uncomfortable with outparty neighbors	Survey 4	-0.161	0.042	-0.243, -0.079
Upset if child marries outparty	Survey 4	-0.112	0.043	-0.196, -0.028
Uncomfortable with outparty friends	Survey 4	-0.146	0.042	-0.228, -0.064
Index: Social distance from outparty	Survey 4	-0.150	0.034	-0.217, -0.083

Table A4: Testing for Divergence (Figure 4)

Outcome	Study	Effect	SE	CI
Inparty	Survey 1	0.052	0.061	-0.068, 0.172
Inparty	Survey 3	-0.016	0.026	-0.067, 0.035
Inparty	Pooled	-0.005	0.024	-0.052, 0.042
Outparty	Survey 1	-0.003	0.067	-0.134, 0.128
Outparty	Survey 3	-0.028	0.043	-0.112, 0.056
Outparty	Pooled	-0.021	0.036	-0.092, 0.05

Table A5: Testing for Desensitization (Figure 5a)

Outcome	Study	Effect	SE	CI
(Cell 3 - Cell 1)	Survey 1	-0.698	0.044	-0.784, -0.612
(Cell 4 - Cell 2)	Survey 1	-0.569	0.054	-0.675, -0.463
(Cell 4 - Cell 2) - (Cell 3 - Cell 1)	Survey 1	0.129	0.070	-0.008, 0.266

Table A6: Testing for Desensitization (Figure 5b)

Outcome	Study	Effect	SE	CI
(Cell 3 - Cell 5)	Survey 3	-0.289	0.039	-0.365, -0.213
(Cell 4 - Cell 6)	Survey 3	-0.223	0.044	-0.309, -0.137
(Cell 4 - Cell 6) - (Cell 3 - Cell 5)	Survey 3	0.066	0.059	-0.05, 0.182

Table A7: Testing for Desensitization (Figure 5c)

Outcome	Study	Effect	SE	CI
(Cell 5 - Cell 1)	Survey 3	-0.148	0.037	-0.221, -0.075
(Cell 6 - Cell 2)	Survey 3	-0.269	0.042	-0.351, -0.187
(Cell 6 - Cell 2) - (Cell 5 - Cell 1)	Survey 3	-0.121	0.056	-0.231, -0.011

Table A8: Testing for Greater Receptivity to Party Cues (Figure 6)

Outcome	Study	Effect	SE	CI
Party Issue Following	Survey 1	0.011	0.024	-0.036, 0.058
Party Issue Following	Survey 3	-0.021	0.027	-0.074, 0.032
Party Issue Following	Pooled	-0.003	0.018	-0.038, 0.032

Table A9: Effect of Bipartisan (instead of Partisan) MC Behavior on Approval of MC (Figure 7)

Outcome	Study	Effect	SE	CI
Effect in High Affpol Condition	Survey 2	0.082	0.050	-0.016, 0.18
Effect in Low Affpol Condition	Survey 2	-0.072	0.056	-0.182, 0.038
Effect in Low Affpol Condition Minus Effect in High Affpol Condition	Survey 2	-0.154	0.075	-0.301, -0.007

Table A10: Testing for Undermined Support for Democratic Norms (Figure 8)

Outcome	Study	Effect	SE	CI
Index: Opposition to Norms	Pooled	-0.037	0.025	-0.086, 0.012
Index: Opposition to Norms	Survey 1	-0.067	0.048	-0.161, 0.027
Index: Opposition to Norms	Survey 2	-0.013	0.039	-0.089, 0.063
Index: Opposition to Norms	Survey 4	-0.042	0.043	-0.126, 0.042
Index: Violence Justification	Survey 4	-0.015	0.043	-0.099, 0.069
Vignette: Do ∗not∗investigate inparty donor in example article.	Pooled	-0.104	0.062	-0.226, 0.018
Vignette: Do ∗not∗investigate inparty donor in example article.	Survey 1	-0.111	0.082	-0.272, 0.05
Vignette: Do ∗not∗investigate inparty donor in example article.	Survey 2	-0.094	0.095	-0.28, 0.092
Vignette: Do ∗not∗use tear gas on inparty protestors.	Survey 1	-0.017	0.083	-0.18, 0.146
Vignette: Inparty legislature should override election of outparty candidate.	Pooled	0.023	0.039	-0.053, 0.099
Vignette: Inparty legislature should override election of outparty candidate.	Survey 2	0.247	0.098	0.055, 0.439
Vignette: Inparty legislature should override election of outparty candidate.	Survey 4	-0.019	0.042	-0.101, 0.063
Vignette: Investigate outparty donor in example article.	Pooled	0.037	0.040	-0.041, 0.115
Vignette: Investigate outparty donor in example article.	Survey 2	0.161	0.097	-0.029, 0.351
Vignette: Investigate outparty donor in example article.	Survey 4	0.012	0.043	-0.072, 0.096
Vignette: News website should ∗not∗post example article criticizing inparty.	Pooled	0.001	0.066	-0.128, 0.13
Vignette: News website should ∗not∗post example article criticizing inparty.	Survey 1	-0.103	0.087	-0.274, 0.068
Vignette: News website should ∗not∗post example article criticizing inparty.	Survey 2	0.143	0.102	-0.057, 0.343
Vignette: News website should post example article criticizing outparty.	Survey 2	0.073	0.095	-0.113, 0.259
Vignette: Outparty legislature should ∗not∗override election of inparty candidate.	Survey 2	-0.138	0.101	-0.336, 0.06

Table A11: Testing for Increased Bias in Perceptions of Objective Conditions (Figure 9)

Outcome	Party	Study	Effect	SE	CI
Perception of ∗fewer∗COVID deaths	Republicans	Survey 4	0.064	0.067	-0.067, 0.195
Perception of ∗greater∗COVID deaths	Democrats	Survey 4	-0.043	0.057	-0.155, 0.069
Perception of ∗higher∗unemployment	Democrats	Survey 4	-0.113	0.057	-0.225, -0.001
Perception of ∗lower∗unemployment	Republicans	Survey 4	0.002	0.068	-0.131, 0.135

Table A12: Survey 5: Manipulation Checks (Figure 10)

Outcome	Manipulation	Effect	SE	CI
DV = Feeling thermometer: Outpartisans	Ahler and Sood (2018)	0.146	0.062	0.024, 0.268
DV = Feeling thermometer: Outpartisans	Lees and Cikara (2020)	0.226	0.061	0.106, 0.346
DV = Feeling thermometer: Outpartisans	Levendusky (2020)	0.255	0.063	0.132, 0.378
DV = Feeling thermometer: Inpartisans	Ahler and Sood (2018)	-0.147	0.059	-0.263, -0.031
DV = Feeling thermometer: Inpartisans	Lees and Cikara (2020)	-0.080	0.057	-0.192, 0.032
DV = Feeling thermometer: Inpartisans	Levendusky (2020)	-0.043	0.058	-0.157, 0.071
DV = Affective polarization (mass public) (Feeling thermometer difference)	Ahler and Sood (2018)	-0.204	0.062	-0.326, -0.082
DV = Affective polarization (mass public) (Feeling thermometer difference)	Lees and Cikara (2020)	-0.230	0.061	-0.35, -0.11
DV = Affective polarization (mass public) (Feeling thermometer difference)	Levendusky (2020)	-0.232	0.061	-0.352, -0.112
DV = Affective polarization (towards elites) (Feeling thermometer difference)	Ahler and Sood (2018)	-0.187	0.062	-0.309, -0.065
DV = Affective polarization (towards elites) (Feeling thermometer difference)	Lees and Cikara (2020)	-0.122	0.061	-0.242, -0.002
DV = Affective polarization (towards elites) (Feeling thermometer difference)	Levendusky (2020)	-0.151	0.062	-0.273, -0.029

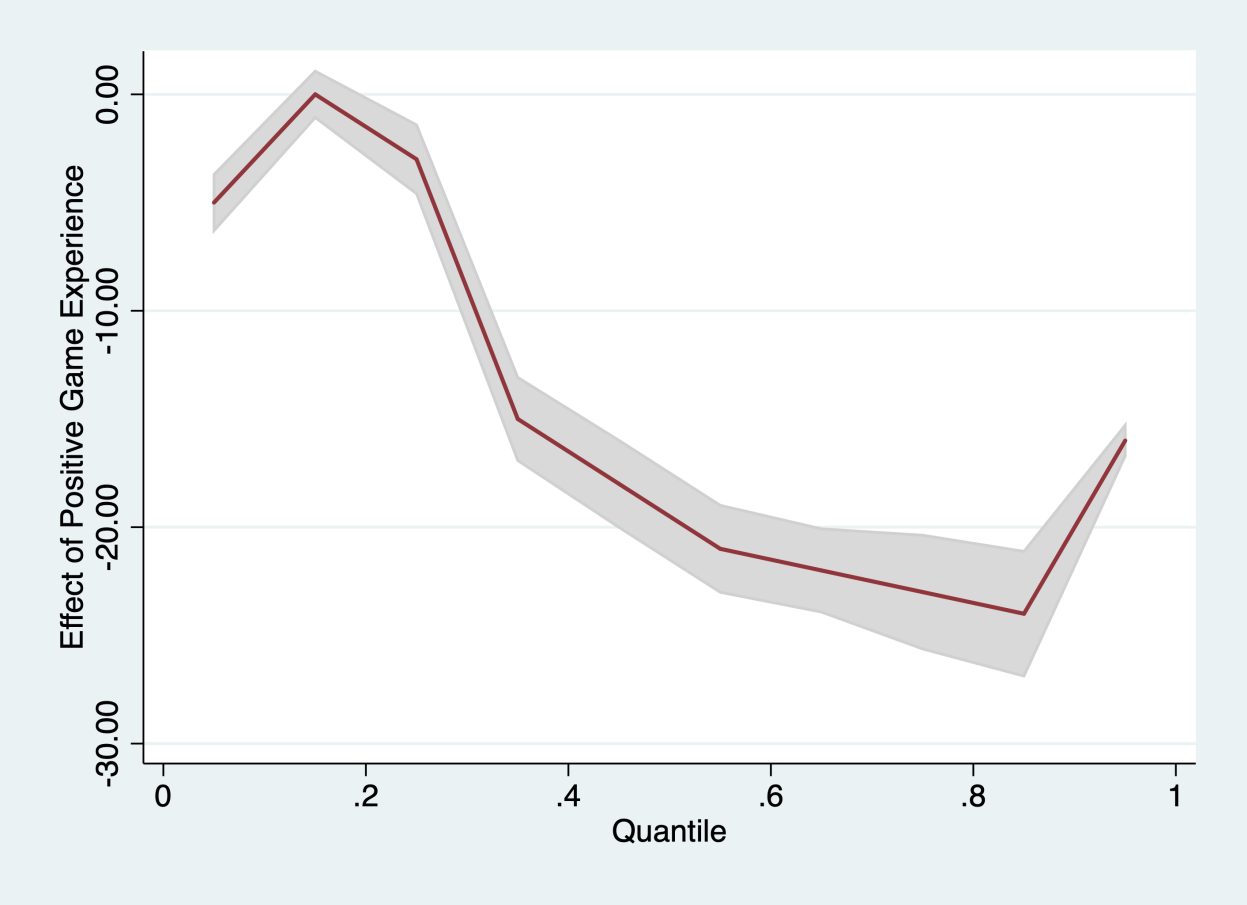
Table A13: Survey 5 Results (Figure 11)

Outcome	Effect	SE	CI
Justified for inparty to harass outparty citizens	-0.050	0.051	-0.15, 0.05
Justified for inparty to threaten outparty leaders	-0.048	0.051	-0.148, 0.052
Justified for inparty to use violence if lose 2024 election	-0.018	0.050	-0.116, 0.08
Justified for inparty to use violence to advance goals today	-0.037	0.051	-0.137, 0.063
Index: Violence Justification	-0.041	0.051	-0.141, 0.059
Ban mail in voting in outparty areas.	0.009	0.050	-0.089, 0.107
Reduce polling places in outparty geographies.	-0.017	0.051	-0.117, 0.083
Criminally investigate journalists who accuse inparty politicians of misconduct without naming sources.	-0.005	0.051	-0.105, 0.095
Inparty governor should use executive orders and ignore outparty legislature.	0.006	0.050	-0.092, 0.104
Inparty governor should ignore outparty-appointed judges.	0.006	0.051	-0.094, 0.106
Inparty should not compromise with outparty.	-0.033	0.051	-0.133, 0.067
Vignette: Inparty legislature should override election of outparty candidate.	-0.016	0.050	-0.114, 0.082
Index: Opposition to Norms	-0.009	0.050	-0.107, 0.089
Uncomfortable with outparty neighbors	-0.117	0.050	-0.215, -0.019
Outpartisan Unattractive	0.060	0.071	-0.079, 0.199
Outpartisan Should Not Be Hired	0.151	0.071	0.012, 0.29

B.1 Quantile Regression Results

An anonymous reviewer suggested we present quantile regressions showing the effect of the trust game experience condition across quantiles of affective polarization. Note that this analysis was not pre-registered.

Figure A7: Quantile Regression



Notes: Estimates from Surveys 1-4 with survey fixed effects. This analysis was not pre-registered.

Appendix Figure A7 presents the results of such a quantile regression, pooled across Surveys 1-4 with survey fixed effects. We find that across most quantiles, the positive trust game successfully reduces affective polarization relative to the negative trust game experience. The one exception is among people already low on affective polarization (e.g., below the 20th percentile, where the

average affective polarization is a feeling thermometer difference of 2 degrees or less), the positive trust game does not further reduce affective polarization among this group. However, even those at the very high end of the distribution are seeing their affective polarization be substantially reduced. In other words, it appears that the trust game manipulation cannot move the minority of people who are essentially not affectively polarized at all to below 0 on affective polarization (i.e., it does not cause people who already feel similarly about the parties to *like* the other party *more* than their own), but for the remainder of the population, including those who are most affectively polarized, do exhibit large effects.

C Surveys

C.1 Survey Contents and Order

Table B1: Survey Contents and Order, by Survey

	Survey 1	**Survey 2**	**Survey 3**	**Survey 4**	**Survey 5**
Demographics Battery	1	1	1	1	1
Asked Own Views on Issues Pre-Treatment	2		2		
Trust Game	3	2	3	2	2$^{\wedge}$
Manipulation Check: Treated Fairly	4	3	4	3	3*
Manipulation Check: Mass Feeling Thermometers	5	4	5	4	4
Manipulation Check: Salience of Partisan Identity	6				
Manipulation Check: Elite Feeling Thermometers				5	5
Social Distance Items				6	6*
Shown MC Votes & Asked MC DVs	7		6		
Norms Index	8	5		7†	7 †
Asked Own Views on Issues Post-Treatment	9		7		
Suppression Vignette	10‡ (In-party only)				
Corruption Tolerance Vignette	11‡ (In-party only)	6‡ (Party randomized)		8† (Out-party only)	
Antilocution Vignette	12‡ (In-party only)	8‡ (Party randomized)			
Election Override Vignette		7‡ (Party randomized)		9† (In-party only)	8† (In-party only)
Bipartisanship Vignette		9			
Violence Justification Index				10	9†
Perceptions of Objective Conditions				11	
Final Demographics	13	10	8	12	10

‡: Respondents were randomized to just one of these vignettes.
$\wedge$: Respondents were randomly assigned to the negative trust game, control, or one of the additional treatments designed to reduce affective polarization.
*: This was only asked for the negative trust game condition.
*: Respondents were also asked the specific social distance vignettes on outpartisan attractiveness or outpartisan worker. Respondents were randomized to just one vignette.
†: The order of these items was randomized.

C.2 Question Text

C.2.1 Demographics Battery

- What is your gender?
  - Male
  - Female
  - Other
- What is your race and ethnicity? Select all that apply.
  - White/Caucasian
  - Hispanic/Latino
  - African American
  - Asian
  - Native American
  - Pacific Islander
  - Other
- What is your combined annual household income?
  - Less than $30,000
  - $30,000 - 39,999
  - $40,000 - 49,999
  - $50,000 - 59,999
  - $60,000 - 69,999
  - $70,000 - 79,999

- $80,000 - 89,999
- $90,000 - 99,999
- $100,000 or more
- Prefer not to say

C.2.2 Own Views on Issues Pre-Treatment

Congress considered many important laws recently.

If you were in Congress would you vote FOR or AGAINST each of the following? (Options: FOR, AGAINST)

- **Working Families Flexibility Act of 2017.** Allows employers to give employees who worked overtime paid time off instead of only overtime pay.
- **Reducing Regulatory Burdens Act of 2017.** Allows pesticides to be sprayed near water sources without obtaining a permit.
- **Thin Blue Line Act.** Allows the death penalty in the case of a murder or attempted murder of police officers, correctional officers, firefighters, or other first responders.
- **Save Local Businesses Act.** If an employee working for a company through a 'temp' agency is injured, only the temp agency is responsible and not the company directing the worker day-to-day.
- **Kate's Law.** Increases criminal penalties for unauthorized immigrants who re-enter the United States after having been deported.
- **Promoting Cross Border Energy Infrastructure Act.** Allows oil and natural gas pipelines that cross into Canada or Mexico to be built without the President's permission.

- **Countering America's Adversaries Through Sanctions Act.** Places additional sanctions on Iran, Russia, and North Korea, as well as individuals who conduct business with these countries.

- **Department of Veterans Affairs Accountability and Whistleblower Protection Act of 2017.** Authorizes the Secretary of Veterans Affairs to demote, suspend, or fire senior Veterans Affairs employees for performance or misconduct, but forbids retaliation against whistleblowers.

- **Financial CHOICE Act of 2017.** Allows banks of sufficient size to take additional risk, and limits the power of the Consumer Financial Protection Bureau to investigate banks.

- **No Sanctuary for Criminals Act.** Prohibits giving federal grants to cities with "sanctuary" policies, policies cities enact to limit their cooperation with federal immigration law enforcement.

- **Ozone Standards Implementation Act of 2017.** Delays the implementation of a rule that would have reduced ozone pollution, allowing previous levels of pollution until 2026.

- **Tax Cuts and Jobs Act.** Reduces corporate taxes from 35% to 21% permanently. Temporarily reduces individual income taxes, with larger reductions for wealthier individuals. Increases the federal budget deficit by $1 trillion.

- **Sportsmen's Heritage and Recreational Enhancement (SHARE) Act of 2015.** Allows individuals to fish and hunt on federal lands without a license, unless the lands are closed for conservation, public safety, or national security.

- **Ozone Standards Implementation Act of 2017.** Delays the implementation of a rule that would have reduced ozone pollution, allowing previous levels of pollution until 2026.

- **No Taxpayer Funding for Abortion and Abortion Insurance Full Disclosure Act of 2017.** Prohibits the use of any federal funds for health insurance that provides abortion services.

- **Veterans 2nd Amendment Protection Act.** Allows any veteran deemed mentally incompetent to buy and own firearms and ammunition, unless a judge deems them dangerous.

- **Prohibits Use of Funds for Discrimination Based on Sexual Orientation or Gender Identity.** Prohibits the government from doing business with companies that discriminate against individuals based on sexual orientation or gender identity.

- **American Health Care Act of 2017.** Repeals "Obamacare": 1) Allows states to allow insurance companies to charge individuals more for insurance if they have a pre-existing condition. 2) Removes the requirement that Americans must carry health insurance. 3) Reduces amount given to low-income Americans to help them purchase health insurance.

C.2.3 Trust Game

Trust Game Instructions.

For our next study, we are going to ask you to play games with other survey respondents.

You will participate in several economic tasks called "games" over the next few minute.

You will be assigned to a different partner (someone else completing this survey) for each game.

You will receive some basic demographic information on each partner, but you will not find out who this person is, nor will he or she find out who you are (not now, nor after the survey is over).

You will work with money for each game. We will pay you an amount based on your final total.

(page break)

Instructions

This game is played by pairs of individuals. Each pair is made up of a Player 1 and a Player 2.

Game steps

We will give $10 to each Player 1. Player 1 then has the opportunity to give a portion of his or her $10 to Player 2. Player 1 could give some, all, or none of the $10. Whatever amount Player 1 decides to give to Player 2 will be tripled before it is passed on to Player 2. Player 2 then has the option of returning any portion of this tripled amount to Player 1. Each Player has 20 seconds to act.

Payment

Player 1 receives whatever he or she kept from their original $10, plus anything returned to him or her by Player 2. Player 2 receives whatever was given to him or her by Player 1, tripled, but then minus whatever they returned to Player 1. Note: We will multiply the final totals by 0.05 and give you a bonus for this survey of that amount. For example, if you win $20, we would pay you a bonus of $20 * 0.05 = $1. Please pay careful attention to these instructions. We will ask practice questions to ensure you understand.

(page break)

We will now run through 3 examples to show you how the game might be played.

Example 1 As always, Player 1 starts with $10. Imagine that Player 1 then gives $4 to Player 2. We triple this amount, so Player 2 gets $12 (3 times $4 equals $12). At this point, Player 1 has $6 and Player 2 has $12. Then Player 2 has to decide whether to give anything back to Player 1, and if so, how much. Suppose Player 2 decides to return $3 to Player 1. At the end of the game Player 1 will have $9 and Player 2 will have $9.

Example 2 Imagine that Player 1 gives all $10 to Player 2. We triple this amount, so Player 2 gets $30 (3 times $10 equals $30). At this point, Player 1 has $0 and Player 2 has $30. Then Player 2 has to decide whether to give anything back to Player 1, and if so, how much. Suppose Player 2 decides to return $15 to Player 1. At the end of the game Player 1 will have $15 and Player 2 will

have $15.

Example 3 Imagine that Player 1 gives $3 to Player 2. We triple this amount, so Player 2 gets $9 (3 times $3 equals $9). At this point, Player 1 has $7 and Player 2 has $9. Then Player 2 has to decide whether to give anything back to Player 1, and if so, how much. Suppose Player 2 decides to return $0 to Player 1. At the end of the game Player 1 will have $7 and Player 2 will have $9.

(page break)

Practice questions

Question 1: Player 1 starts with $10. Suppose that Player 1 gives $7 to Player 2. How much money will Player 2 get?

- $7
- $14
- $21

Question 2: After getting the money, what can Player 2 do with the money?

- Keep all the money
- Give some of the money to Player 1
- Give all of the money to Player 1
- All of the above

If correct: Correct!

If incorrect: Please try again. You must answer all questions correctly before you can continue. Player 1 starts with $10. Suppose that Player 1 gives $2 to Player 2. How much money will Player 2 get?

- $0

- $2
- $6

Question 2: After getting the money, what can Player 2 do with the money?

- Keep all the money
- Give some of the money to Player 1
- Give all of the money to Player 1
- All of the above

Trust Game.

You have been randomly assigned to play as Player 2. You will play as Player 2 for three rounds.

Each Player 1 will see the following information about you. They will use it to decide how trustworthy you are.

Age: [respondent's own age] Gender: [respondent's own gender] Income: [respondent's own income] Partisanship: [respondent's own party]

Round 1/2/3. (This is repeated for 3 rounds.)

You are Player 2.

Player 1 can give you some, all, or none of the $10. We will triple any amount Player 1 allocates to you. You are under no obligation to give anything back.

Information about who you are playing with (Player 1):

Age: [randomized age] Gender: [randomized gender] Income: [randomized income group] Partisanship: [respondent's out party]

Please wait while Player 1 decides your allocation.

(timer from 20 seconds counts downs, advances after about 5 seconds)

(page break)

If in Negative Experience group: *Results.* Player 1 allocated you $0. We are unable to triple this amount.

If in Positive Experience group:

Results. Player 1 allocated you $8. We have tripled this to $24. You can now return some, all or none of this money to Player 1. Put the number of dollars you wish to keep in the box labeled "Player 2." Put the dollars you wish to go to Player 1 in the box labeled "Player 1."

You gave $[amount given back] back to Player 1. Which factors, if any, were part of your decision making process?

- Age
- Gender
- Income
- Partisanship
- Something else

Player 1's reason for their allocation to you: your partisanship (all rounds), your income (Round 2)

Game Summary.

Round 1: Earnings: $[amount] Factors about you that Player 1 used when deciding how much money to give: Political Party.

Round 2: Earnings $[amount] Factors about you that Player 1 used when deciding how much money to give: Political Party, Income.

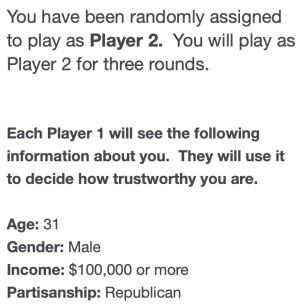
Round 3: Earnings $[amount] Factors about you that Player 1 used when deciding how much money to give: Political Party.

Your total earnings: $[amount]

Trust Game Screenshots

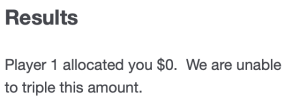
Figure B1: Steps in Trust Game: Screenshots

(a) Step 1



(b) Step 2

(c) Step 3



(d) Step 4

C.2.4 Manipulation Checks

Would you say that you were treated fairly or unfairly when playing the game?

- Very fairly
- Fairly
- Unfairly
- Very unfairly

We'd like you to rate how you feel towards some groups on a scale of 0 to 100. Zero means very unfavorable and 100 means very favorable. Fifty means you do not feel favorable or unfavorable. How would you rate your feeling toward each of the following?

- People who are [out party]s
- People who are [in party]s
- *Survey 4 only.* [out party] Politicians and Elected Officials
- *Survey 4 only.* [in party] Politicians and Elected Officials

- *Surveys 1-3 only.* People who are White
- *Surveys 1-3 only.* People who are Black
- *Surveys 1-3 only.* People who are Poor
- People who are Young
- People who are Old

How important is being a [in party] to how you see yourself as a person?

- Extremely important
- Very important
- Somewhat important
- Not very important
- Not at all important

C.2.5 Social Distance Items

How comfortable would you be to talk about politics with people who are [out party]s?

- Not at all comfortable
- Not too comfortable
- Somewhat comfortable
- Extremely comfortable

How comfortable are you having close personal friends who are [out party]s?

- Not at all comfortable

- Not too comfortable
- Somewhat comfortable
- Extremely comfortable

How comfortable are you having neighbors on your street who are [out party]s?

- Not at all comfortable
- Not too comfortable
- Somewhat comfortable
- Extremely comfortable

Suppose a son or daughter of yours was getting married. How would you feel if he or she married a supporter of the [out party]s?

- Not at all upset
- Not too upset
- Somewhat upset
- Extremely upset

C.2.6 Shown MC Votes and Asked MC DVs

Welcome to Study 3 of 3 in this survey. In this part, we will show you information about your Representative in Congress and ask for your reaction to it.

Your Representative in Congress is [Representative Name and Party].

From the bills in Congress we asked you about before, we have chosen three at random. We want to show you how [MC Name] actually voted on those bills.

- [*Bill title 1*]. Bill title description: [MC Name] Voted [Yes/No].
- [*Bill title 2*]. Bill title description: [MC Name] Voted [Yes/No].
- [*Bill title 3*]. Bill title description: [MC Name] Voted [Yes/No].

If the 2020 Congressional election were held today, who would you vote for?

- [MC Name and Party]
- The [Opposite party of MC] that runs against them
- For a third party
- Other
- I would not vote

If Third party, Other, or Would not vote is selected: If you had to choose in the 2020 election between [MC Name] and the [Opposite party of MC] who runs against them, who would you lean towards?

- [MC Name]
- The [Opposite party of MC] that runs against them
- Completely undecided

Do you approve or disapprove of the way [MC Name and Party] is handling their job as your representative in Congress?

- Strongly approve
- Approve
- Somewhat approve

- Neither approve nor disapprove
- Somewhat disapprove
- Disapprove
- Strongly disapprove

How favorable is your overall opinion of [MC Name and Party]?

- Extremely favorable
- Moderately favorable
- Slightly favorable
- Neither favorable nor unfavorable
- Slightly unfavorable
- Moderately unfavorable
- Extremely unfavorable

C.2.7 Social Distance Index

How comfortable would you be to talk about politics with people who are [outparty]s?

- Not at all comfortable
- Not too comfortable
- Somewhat comfortable
- Extremely comfortable

How comfortable are you having close personal friends who are [outparty]s?

- Not at all comfortable
- Not too comfortable
- Somewhat comfortable
- Extremely comfortable

How comfortable are you having neighbors on your street who are [outparty]s?

- Not at all comfortable
- Not too comfortable
- Somewhat comfortable
- Extremely comfortable

Suppose a son or daughter of yours was getting married. How would you feel if he or she married a supporter of the [outparty]s?

- Not at all upset
- Not too upset
- Somewhat upset
- Extremely upset

C.2.8 Norms Index

Do you agree or disagree with the statements below? (Options: Strongly agree, Agree, Somewhat agree, Neither agree nor disagree, Somewhat disagree, Disagree, Strongly disagree)

- When possible, [in party] politicians should try to compromise with [out party] politicians to get things done.

- *Study 1 only.* When determining the boundaries for Congressional districts, [in party] politicians should try to maximize the number of [in party]s elected.
- If a [in party] governor of a state can't get cooperation from [out party] legislators to pass new laws, the [in party] governor should issue executive orders on their own to accomplish their priorities.
- [in party] elected officials should sometimes consider ignoring court decisions when the judges are appointed by [out party] presidents.
- If a journalist accuses a [in party] politician of misconduct without naming their sources, the journalist should be criminally investigated.
- [in party]s should reduce the number of polling stations in areas that typically supports [out party]s.
- *Study 1 only.* Far-[left (if respondent is Republican) / right (if respondent is Democrat)] rallies should be banned.
- *Study 4 only, not included in norms scale.* Mail-in voting should be banned in areas that typically support [out party]s.

C.2.9 Own Views on Issues Post-Treatment

If you were in Congress would you vote FOR or AGAINST each of the following? (Options: For, Against)

- [*Bill title 1*]. Bill title description.
- [*Bill title 2*]. Bill title description.
- [*Bill title 3*]. Bill title description.

C.2.10 Vignette: Suppression

Please read the news story below.

Police Use Tear Gas on Peaceful Young [*Party*] *Protest*

Riot police broke up a group of Young [Party]s protesting a tuition increase at a college north of New York yesterday, using what appeared to be tear gas.

Students had been told earlier that ”all necessary force” would be used to ensure the protests did not interfere with the administration, in line with a recent court injunction requested and received by 53 of the school's students.

Witnesses reported that 30 police officers deployed to confront the 80 peaceful protesters. After some jostling and use of what appeared to be tear gas by police, the crowd blocking the doors broke and moved away.

Alan Krenshaw, a spokesperson for the Young [Party]s, complained that “The brutality was absolutely designed to chill the movement and literally try to beat and terrorize our right to criticize, to think critically and to act on that criticism.”

Do you agree or disagree with the decision to use tear gas on the [Party] protesters?

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree

- Strongly disagree

The cost of the police response is unknown at this point, but the city can fine the protests any amount up to $10,000. What amount, if any, do you think the city should fine the group of protesters? Enter a number between 0 and 10,000 below. (Do not include a $ sign.)

C.2.11 Corruption Tolerance

Please read the news story below.

> *Donations from Millionaire Businessman to* [*Party*] *Super PACs in Question*
>
> Federal election officials announced this morning that 6 million dollars in donations from Alan Gregory, a retired millionaire, to [Party] Super PACs are now under investigation.
>
> Officials allege that the donations were made through a questionable middle-man over a period of two months. The case is particularly important, as support from Mr. Gregory is credited in the victory of several [Party] state officials in the last election.
>
> Mr. Gregory and [Party] party officials deny the allegations and claim that although the donations were not transparent, they were not illegal.

Do you support the investigation of the businessman?

- Strongly support
- Support
- Somewhat support
- Neither support or oppose
- Somewhat oppose

- Oppose
- Strongly oppose

C.2.12 Antilocution Vignette

Please read the news article below.

> *OPINION: [Party]s Drive Congress to Do Less Than Last Year's Record-Breaking Low*
>
> Led by the [Party]'s refusal to compromise, the current U.S. Congress faces a backlog of unfinished business and sliding approval ratings.
>
> [Party]s must recognize that the country is not behind their obstructionist tactics. We need to move forward without letting crazy members of the opposition hold our nation hostage.
>
> Out of touch [Party]s are causing the gridlock by being unwilling and unable to engage in the normal process of negotiation and compromise. Major pieces of legislation, including a budget agreement and a farm and food-aid policy bill, are sitting on the sidelines.

A large website that posts stories from many different news sources is considering sharing the article you just read. Do you think they should post this article?

- Definitely
- Probably
- Maybe
- Probably not
- Definitely not

C.2.13 Election Override Vignette

Please read the news story below.

> *Local* [*Party*] *Candidate Calls for State Legislature to Decide Election*
>
> After the State Board of Elections announced that [Party] candidate Valerie Johnston lost by a few hundred votes, Johnson is petitioning the [Party]-controlled state legislature to take matters into its own hands and decide the election in her favor. According to the State Board of Elections, Johnston, the [Party] candidate, received 217 fewer votes than the [Other Party] candidate, Stephanie Cole. The result leaves the two candidates less than 0.1% apart after a bitter campaign.
>
> Talking to reporters, Johnston noted "that she had heard many claims that voting machines weren't working correctly in [Party] neighborhoods and that absentee ballots were not correctly handled." State law gives the legislature final say over certifying election outcomes, a provision meant to be used in cases of fraud or widespread errors. Johnson claims the law was intended to allow the state legislature to intervene in cases exactly like hers.
>
> "Whatever process that the state legislature chooses to go down, it's not going to change the overall number of votes," Cole said at a press conference. "In a democracy, we cannot allow a candidate that loses an election to ask her [Party] allies in the state legislature to just overturn election results. In a democracy, we must respect the will of the people."
>
> A spokesperson for the State Board of Elections told reporters that Johnston's complaint was "without basis." Meanwhile, the state legislature is debating whether to vote on the issue and determine the outcome of the election itself.

Do you support or oppose the [Party] state legislature determining the outcome of the election?

- Strongly support
- Support
- Somewhat support
- Neither support or oppose
- Somewhat oppose
- Oppose
- Strongly oppose

C.2.14 Bipartisanship Vignette

Note: This vignette is from Harbridge and Malhotra (2011, Study 2).

Republicans, Bipartisan Condition. Some members of Congress work with members of their own party almost all of the time. Other members work with members of both parties. In a recent year, on key issues identified by the American Conservative Union (ACU), Representative Steve LaTourette (a Republican from Ohio) took the Republican position on about half the votes and the Democratic position on about half the votes.

Republicans, Partisan Condition. Some members of Congress work with members of their own party almost all of the time. Other members work with members of both parties. In a recent year, on key issues identified by the American Conservative Union (ACU), Representative Steve LaTourette (a Republican from Ohio) almost always voted the Republican position.

Democrats, Bipartisan Condition. Some members of Congress work with members of their own party almost all of the time. Other members work with members of both parties. In a recent year, on key issues identified by the Americans for Democratic Action (ADA), Representative Bud Cramer (a Democrat from Alabama) almost always voted the Democratic position.

Democrats, Partisan Condition. Some members of Congress work with members of their own party almost all of the time. Other members work with members of both parties. In a recent year, on key issues identified by the Americans for Democratic Action (ADA), Representative Bud Cramer (a Democrat from Alabama) took the Democratic position on about half the votes and the Republican position on about half the votes.

Based on this information, do you approve or disapprove of the job [Representative LaTourette (R) / Representative Cramer (D)] is doing in Congress?

- Strongly approve
- Somewhat approve
- Neither approve nor disapprove
- Somewhat disagree
- Strongly disagree

C.2.15 Violence Justification Index

How much would it be justified for [in party]s. . . (Options: Not at all, A little, A moderate amount, A lot, A great deal)

- . . . to use violence to advance their political goals these days?
- . . . to use violence if they lose the 2024 election?

When (if ever) do you think it is ok for an ordinary person who is a [in party] . . . (Options: Never, Rarely, Sometimes, Often)

- . . . to harass an ordinary [out party] on the Internet in a way that makes the [out party] person feel unsafe?
- . . . to send physical threats and intimidating messages to [out party] leaders?

C.2.16 Perception of Objective Conditions

Donald Trump has been President of the United States for four years.

What is your best guess about the current unemployment rate in the United States? This is the percent of people who want to work who don't have a job.

- Over 10% unemployed
- 9-10% unemployed
- 8-9% unemployed
- 7-8% unemployed
- 6-7% unemployed
- 5-6% unemployed
- 4-5% unemployed
- Under 4% unemployed

Donald Trump has been President of the United States for four years.

What is your best guess about the number of Americans who have died from COVID-19?

- Fewer than 100,000
- 100,000 - 200,000
- 200,001 - 300,000
- 300,001 - 400,000
- 400,001 - 500,000
- More than 500,000

C.2.17 Final Demographics

What is the highest level of education you have completed?

- Less than High School
- High School / GED
- Some College
- 2-year College Degree
- 4-year College Degree
- Masters Degree
- Doctoral Degree
- Professional Degree (JD, MD)

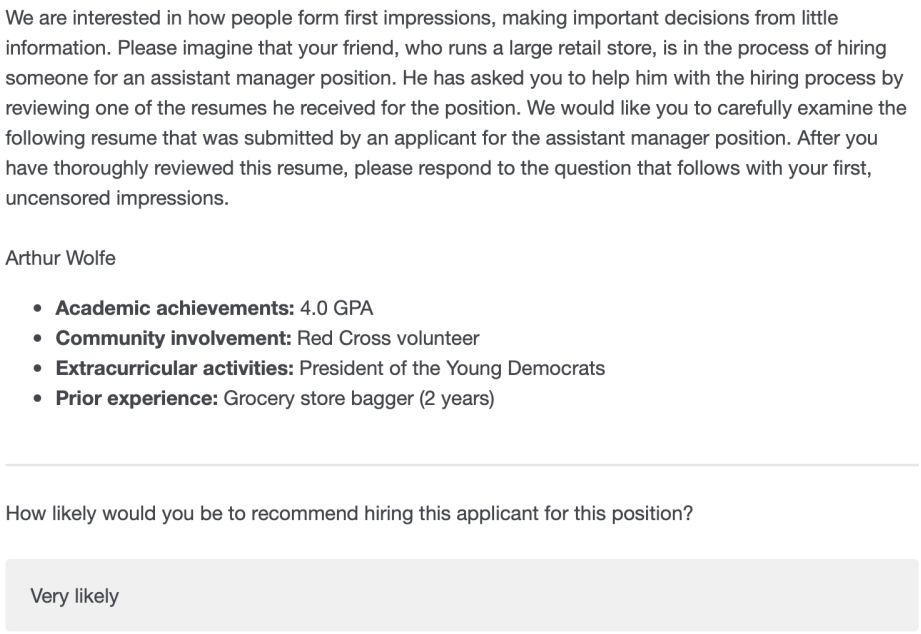
When it comes to politics do you usually think of yourself as...

- Very conservative
- Conservative
- Moderate
- Liberal
- Very liberal
- Don’t know/ None of the above

C.2.18 Job Profile

The displayed job profile was always an outpartisan. No other changes were made to the profile.

Figure B2: Survey 5 Outpartisan Worker Question



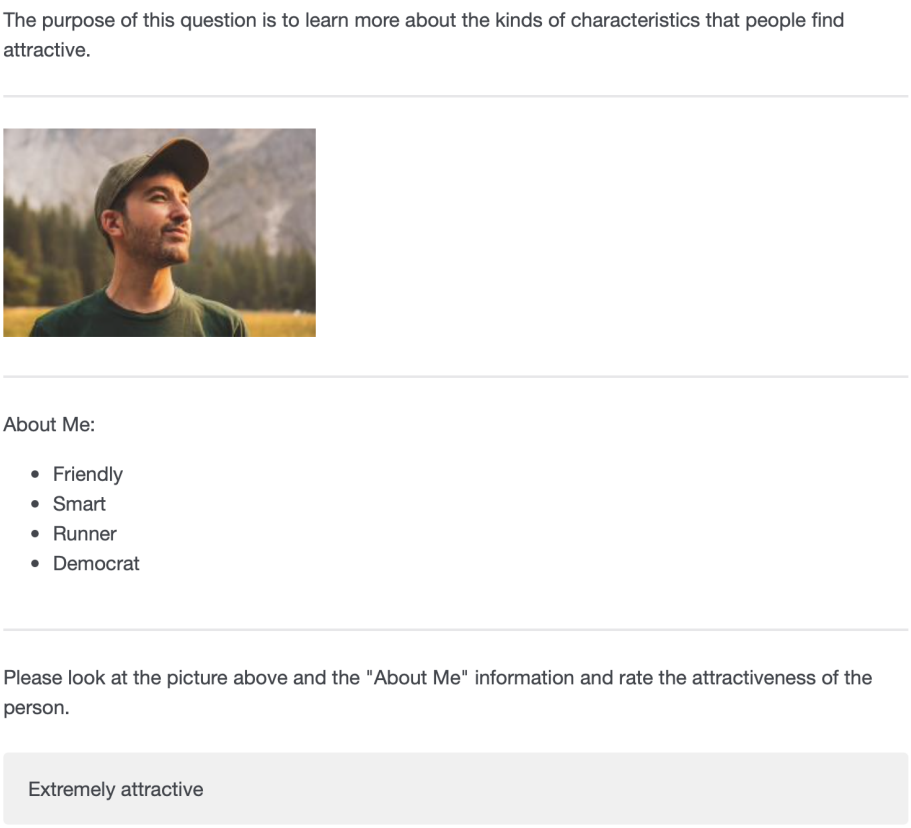
The full question asked: How likely would you be to recommend hiring this applicant for this position?

- Very likely
- Somewhat likely
- Neither likely nor unlikely
- Somewhat unlikely
- Very unlikely

C.2.19 Attractiveness

The profile shown was always an opposite-gender (with corresponding photo) outpartisan. No other changes were made to the profile.

Figure B3: Survey 5 Outpartisan Attractiveness Question



The full question asked: Please look at the picture above and the “About Me” information and rate the attractiveness of the person.

- Extremely attractive
- Moderately attractive
- Slightly attractive
- Neither attractive nor unattractive
- Slightly unattractive
- Moderately unattractive
- Extremely unattractive

C.3 Text of Additional Manipulations (Survey 5)

C.3.1 Ahler and Sood (2018) for Democrats

We are interested in how Americans perceive the two main political parties. To keep the survey short, some respondents will see questions about Republicans, and others about Democrats.

Just give us your best guesses to the questions below. At the end of these questions, we will give you the correct answers so you can check how well you did.

What percentage of Republicans do you think are . . . ?

- Age 65+
- Earn over $250k
- Evangelical
- Southern

(page break)

The percentage of Republicans who are Age 65+ is [smaller/larger] than you think. Only 21.3% are Age 65+. (You [overestimated/underestimated] by [CALC]%).

The percentage of Republicans who earn over $250k is [smaller/larger] than you think. Only 2.2% earn over $250k. (You [overestimated/underestimated] by [CALC]%).

The percentage of Republicans who are Evangelical is [smaller/larger] than you think. Only 34.3% are Evangelical. (You [overestimated/underestimated] by [CALC]%).

The percentage of Republicans who are Southern is [smaller/larger] than you think. Only 35.7% are Southern. (You [overestimated/underestimated] by [CALC]%).

C.3.2 Ahler and Sood (2018) for Republicans

We are interested in how Americans perceive the two main political parties. To keep the survey short, some respondents will see questions about Republicans, and others about Democrats.

Just give us your best guesses to the questions below. At the end of these questions, we will give you the correct answers so you can check how well you did.

What percentage of Democrats do you think are . . . ?

- Black
- Atheist or agnostic
- Union members
- Gay, lesbian, or bisexual

(page break)

The percentage of Democrats who are Black is [smaller/larger] than you think. Only 23.9% are Black. (You [overestimated/underestimated] by [CALC]%).

The percentage of Democrats who are atheist or agnostic is [smaller/larger] than you think. Only 8.7% are atheist or agnostic. (You [overestimated/underestimated] by [CALC]%).

The percentage of Democrats who are union members is [smaller/larger] than you think. Only 10.5% are union members. (You [overestimated/underestimated] by [CALC]%).

The percentage of Democrats who are gay, lesbian, or bisexual is [smaller/larger] than you think. Only 6.3% are gay, lesbian, or bisexual. (You [overestimated/underestimated] by [CALC]%).

C.3.3 Lees and Cikara (2020)

A state [INPARTY] party in control of the state legislature is considering a proposal to change the name given to the state highway. However, currently the state's highway is named after the state's first governor, a [OUTPARTY], who is regarded as the symbolic forbearer of the state [OUTPARTY] party.

How much do you believe a [OUTPARTY] will dislike this action? (0-100 slider)

How much do you believe a [OUTPARTY] opposes this action? (0-100 slider)

How much do you believe a [OUTPARTY] thinks this action is politically unacceptable? (0-100 slider)

In the scenario you just read, which political party was taking action?

- Democratic Party
- Republican Party
- Party A
- Party B
- Not Sure

(page break)

On the previous scenario you estimated [NUM1], [NUM2], and [NUM3] when asked how much a [OUTPARTY] would dislike, oppose, and find politically unacceptable the behavior in the scenario.

However, when we conducted a survey including this scenario with a nationally representative sample, the actual responses of the average [OUTPARTY], when reading the exact same scenario, were 64, 63, and 65.

Moreover, when other [INPARTY]s in the same sample read a similar scenario where [OUTPARTY]s were taking the same action, the actual responses of the average [INPARTY] were 67, 69, and 71.

C.3.4 Levendusky (2020)

Although you are a [INPARTY], you likely know people who are [OUTPARTY]s. Think about one such [OUTPARTY] 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.

Please think about the answer carefully, and then write it in the space below in as much detail as you would like. To give you time to reflect on your answer, the "Continue" button will not appear for a few moments.

C.4 Survey Demographics and Representativeness

Table B2: Comparing survey demographics with the 2019 CCES

	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5	2019 CCES
Party ID - Strong Democrat	23.5%	22.6%	29.7%	30.0%	34.3%	30.6%
Party ID - Not very strong Democrat	14.3%	17.2%	17.1%	17.3%	18.1%	12.9%
Party ID - Lean Democrat	9.6%	12.6%	6.2%	11.5%	10.6%	11.4%
Party ID - Independent*	0%*	0%*	0%*	0%*	0%*	0%*
Party ID - Lean Republican	11.2%	10.9%	6.6%	8.8%	7.6%	10.6%
Party ID - Not very strong Republican	16.0%	16.2%	16.1%	13.5%	11.3%	10.6%
Party ID - Strong Republican	25.6%	20.6%	24.3%	18.8%	18.2%	23.9%
Ideology - Very liberal	8.2%	10.5%	10.8%	11.3%	15.5%	15.5%
Ideology - Liberal	16.2%	17.8%	19.8%	20.4%	21.3%	18.8%
Ideology - Moderate	34.9%	35.8%	31.6%	37.3%	34.1%	28.7%
Ideology - Conservative	27.5%	23.5%	24.8%	20.0%	17.9%	20.4%
Ideology - Very conservative	13.2%	12.5%	13.0%	10.9%	11.2%	16.6%
Age	61.2	45.7	52.7	51.1	50.3	49.8
% Female	59.4%	52.1%	50.2%	54.8%	50.8%	52.0%
% White	90.1%	82.0%	85.1%	71.6%	68.7%	69.7%
% Hispanic or Latinx	1.7%	6.4%	4.0%	12.2%	10.1%	11.6%
% Black	4.8%	7.0%	7.8%	14.6%	11.3%	12.2%
% with income less than $50,000	33.9%	43.5%	31.4%	44.0%	29.8%	41.9%
% with income greater than $100,000	23.4%	18.2%	28.2%	21.6%	30.4%	16.9%
% with a 4-year college degree	52.1%	41.5%	57.4%	42.9%	62.7%	32.7%

Note: Data comes from the 2019 Cooperative Congressional Election Study with common weights.
** Recall that we do not include pure independents in the surveys. To match our sampling frame, we also drop respondents to the CCES who stated they were Independents after being asked a lean question or were not sure of their party identification.*

In addition, on Survey 5, we asked a political knowledge question to benchmark the political knowledge of samples recruited from our survey vendor with our particular approach, exclusions rule, etc. We picked a question from the 2019 Pew Research Center's American Trends Panel, Wave 57. The question asked "As you may know, presidents are chosen not by direct popular vote, but by the electoral college in which each state casts electoral votes. What determines the number of electoral votes a state has?" The response options were "The number of seats the state has in the U.S. House and Senate" (correct), "The number of voters in the state," "The number of counties in the state," "Each state has the same number of electoral votes," and "Not sure." In survey 5, 37% of respondents got this question correct, versus 39% of respondents to the Pew survey.

C.5 Attrition

If subjects in different treatment conditions drop out of the survey at different rates (differential attrition), this may introduce bias into our estimated treatment effects. We find that there are statistically significant but very small differences between the share of individuals who completed the survey in different treatments in some of our surveys. Moreover, because the rates of survey completion among those assigned to treatment are extremely high and the differences between completion rates by treatment are small, we do not find that our confidence intervals grow particularly meaningfully.

In particular, in Table B3, we report survey completion rates by survey and by treatment condition. In the bottom row, we report a p-value from Pearson's chi-squared test to assess how likely it is that the observed differences between the treatment arms in each survey arose by chance.

Table B3: Survey Completion Rates by Treatment and Survey

	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5
Positive Game	91.3%	94.0%	97.3%	94.1%	
Negative Game	96.1%	96.5%	98.2%	94.5%	97.5%
Control					98.1%
Ahler-Sood					95.2%
Lees-Cikara					96.7%
Outparty Friend					94.8%
p-value	<0.001	0.003	0.051	0.664	0.015

From Table B3, we first observe high rates of survey completion across all treatment arms. However, we do observe some small but statistically significant differential attrition in Surveys 1, 2, and 5.

In order to assess whether the treatment effects we report in the main text are robust to addressing this differential attrition, in Table B4 we report 95% confidence intervals of the effect of each treatment on our measure of affective polarization using Lee (2009) trimming bounds. Overall, we find statistically significant effects on affective polarization, even after using

Table B4: 95% Confidence Intervals of Treatment Effect on Affective Polarization Using Lee (2009) Bounds

	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5
Positive vs. Negative	[-0.530, -0.285]	[-0.613, -0.365]	[-0.487, -0.323]	n/a	
Negative vs. Control					[-0.080, 0.256]
Ahler-Sood vs. Control					[-0.363, 0.029]
Lees-Cikara vs. Control					[-0.357, -0.017]
Outparty Friend vs. Control					[-0.400, -0.016]

trimming bounds. This suggests that the effects on affective polarization are not artifacts of differential attrition.

We do not report trimming bounds for any of the downstream outcomes. Trimming bounds would only increase the width of confidence intervals that we report in the main text. The null effects that we report in the main text would therefore remain null.

We also note that prior to randomization, 48% of participants who started Survey 1 were deemed eligible because they met all criteria (informed consent, identify as a Democrat or a Republican (including as a leaner), pass a pre-treatment attention check, and demonstrate understanding of the trust game), 45% for Survey 2, 56% for Survey 3, and 52% for Survey 4. Ineligible respondents were dropped prior to randomization. Respondents were also ineligible if they had participated in an earlier survey.

D Supplemental Data on Trust Game

As part of a survey for a different project conducted by one of the authors, we conducted supplemental data collection using the Qualtrics Panel. In this survey, $N = 249$ respondents were randomly assigned to either a control group, the negative trust game condition, or the positive trust game condition. (This stands in contrast to the data collection for Surveys 1-4 in the main text, where there was no pure control, and all respondents were assigned to either the negative or positive trust game condition.) We then assessed treatment effects on affective polarization. This survey did not measure downstream political outcomes.

Survey 5 (described in Appendix E) also had a negative trust game condition and a control condition (but no positive trust game condition).

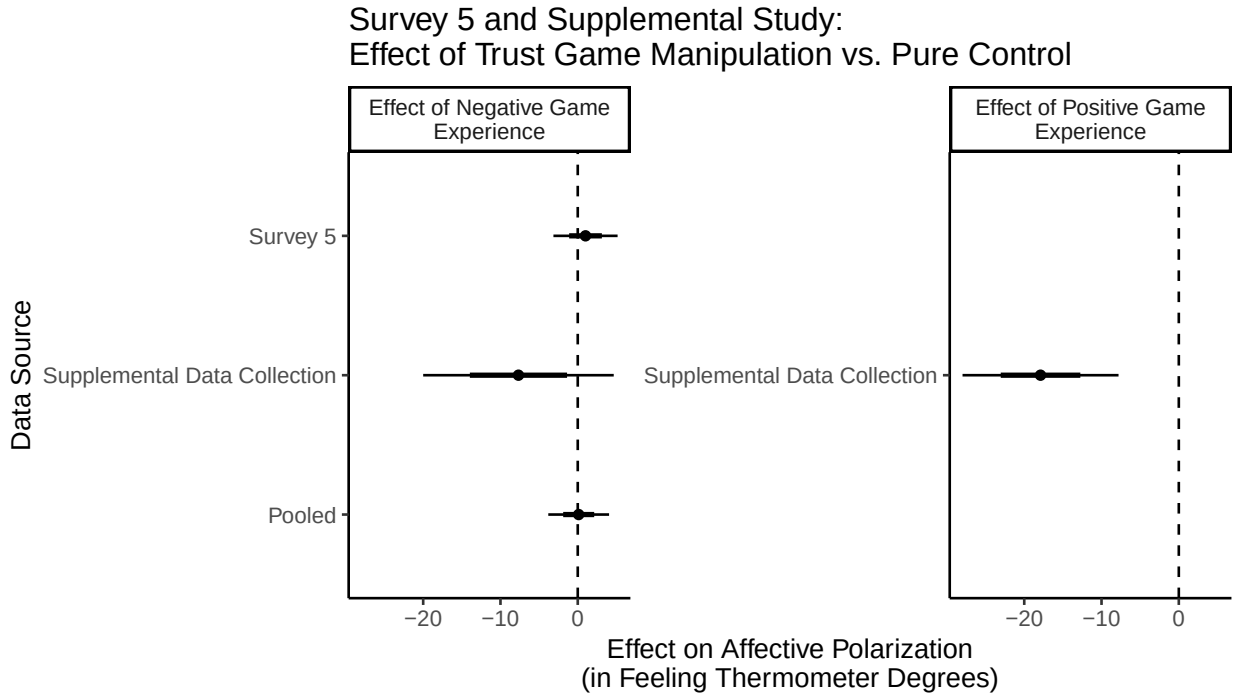
Figure B4 shows the results. Overall, we find that the trust game manipulates affective polarization by *decreasing* affective polarization in the positive game experience condition relative to the negative game experience condition.

First, the left panel of Figure B4 shows that there is a null effect of the negative trust game condition on affective polarization. The first row shows the difference between the level of affective polarization in the control condition versus the negative trust game experience condition in Survey 5; i.e., the effect of the negative trust game condition as compared to control. The second row shows the same effect in the supplemental data collection. The final row pools these two coefficients. The results are null in both datasets. Moreover, when pooling these two coefficients, the average effect of the negative game experience on affective polarization is 0.12 degrees, essentially exactly zero ($SE = 2.01, t = 0.06$).

The second panel of Figure B4 shows that, relative to a control group, the positive trust game condition dramatically decreases affective polarization, by 17.9 points ($SE = 5.1, p < 0.001$). (There was no positive game experience condition in Survey 5.)

These results indicate that one should interpret the differences between the positive and

Figure B4: Supplemental Data on Trust Game: Effect of Positive and Negative Trust Games vs. Control



Notes: Point estimates are surrounded by standard errors (thick lines) and 95% confidence intervals (thin lines). Point estimates are from multivariate regressions controlling for pre-registered covariates (7-point partisanship scale and its absolute value (partisan strength)). Precision-weighting is used to pool across the two surveys.

negative trust game conditions presented in the main text as mainly reflecting the effects of a *decrease* in affective polarization brought on by the positive trust game.

This finding is consistent with the motivation for our research, which raises questions about whether decreasing affective polarization would meaningfully bolster norms or accountability. As compared to the interventions reported in Appendix E, the positive trust game condition is perhaps the most powerful tool for reducing affective polarization available—yet we found it had no salutary downstream consequences for democratic norms or accountability.

There are several potential explanations for why the positive trust game condition so meaningfully decreases affective polarization while the negative trust game condition does not increase it. One possibility is that respondents already expect to be treated poorly by outpartisans,

so the negative trust game provides no new or surprising information; the interaction proceeds exactly as they expect. However, respondents may not expect to be treated well by outpartisans, leading to the reductions in affective polarization observed in the positive trust game. Moreover, respondents in the positive game condition receive an actual financial bonus because of the putative outpartisan's behavior. Future research may wish to investigate these mechanisms.

E Survey 5: Additional Details and Results

Survey 5 was conducted in June 2021 using the online survey vendor Dynata, the same vendor as for surveys 1-4. In order to be eligible to participate, survey respondents needed to meet the same eligibility criteria as Surveys 1-4. In addition, survey respondents were ineligible if they already participated in Surveys 1-4 or were completing the survey on a mobile device. The study was pre-registered with the Center for Open Science.[19] $N = 2,504$ respondents completed the survey. The eligiblity rate was 30% for Survey 5. This survey received IRB approval.

Table B2 presents the demographics of the survey samples relative to the 2019 Cooperative Congressional Election Study. In addition, to assess the political knowledge of the sample, we asked a political knowledge question that had previously been asked by the Pew Research Center.[20] We asked respondents "As you may know, presidents are chosen not by direct popular vote, but by the electoral college in which each state casts electoral votes. What determines the number of electoral votes a state has?" The response options were "The number of seats the state has in the U.S. House and Senate" (correct), "The number of voters in the state," "The number of counties in the state," "Each state has the same number of electoral votes," and "Not sure." In the Pew sample, 39% answered correctly while in the Survey 5 sample, 37% answered correctly. This suggests that the Survey 5 sample was, if anything, slightly less politically knowledgeable than the general population.

After collecting demographics, we randomly assigned respondents to one of four treatment conditions meant to induce variation in affective polarization or to a control group. This follows a similar design as Voelkel et al. (2021). The four treatment conditions (with the full text available in OA Section C.3) were:

- Negative Trust Game. This is the same as in Surveys 1-4.

[19] Pre-analysis plan is available at `https://osf.io/v3qc8/?view_only=eef037e9e32b4a4ea4fed54befd2190f`.

[20] See `https://perma.cc/7V2H-C856`.

- Ahler and Sood (2018). Respondents are asked about their perceptions of the outparty's composition (e.g., What percentage of Republicans do you think are Evangelical?) and then have any misperceptions corrected.

- Lees and Cikara (2020). Respondents are asked about their perceptions of outparty opposition to a policy supported by the inparty and then have any misperceptions of the level of outparty support corrected. They are also shown that Democrats and Republicans both think similarly about this issue.

- Levendusky (2020). An outpartisan friendship is made salient through a writing prompt in which respondents are asked to reflect on a friend, relative, neighbor, or co-worker from the other party who they "like and respect a great deal."

E.1 Results

As shown in Figure 10, the Ahler and Sood (2018), Lees and Cikara (2020), and Levendusky (2020) treatments successfully reduced affective polarization compared to a control group.[21]

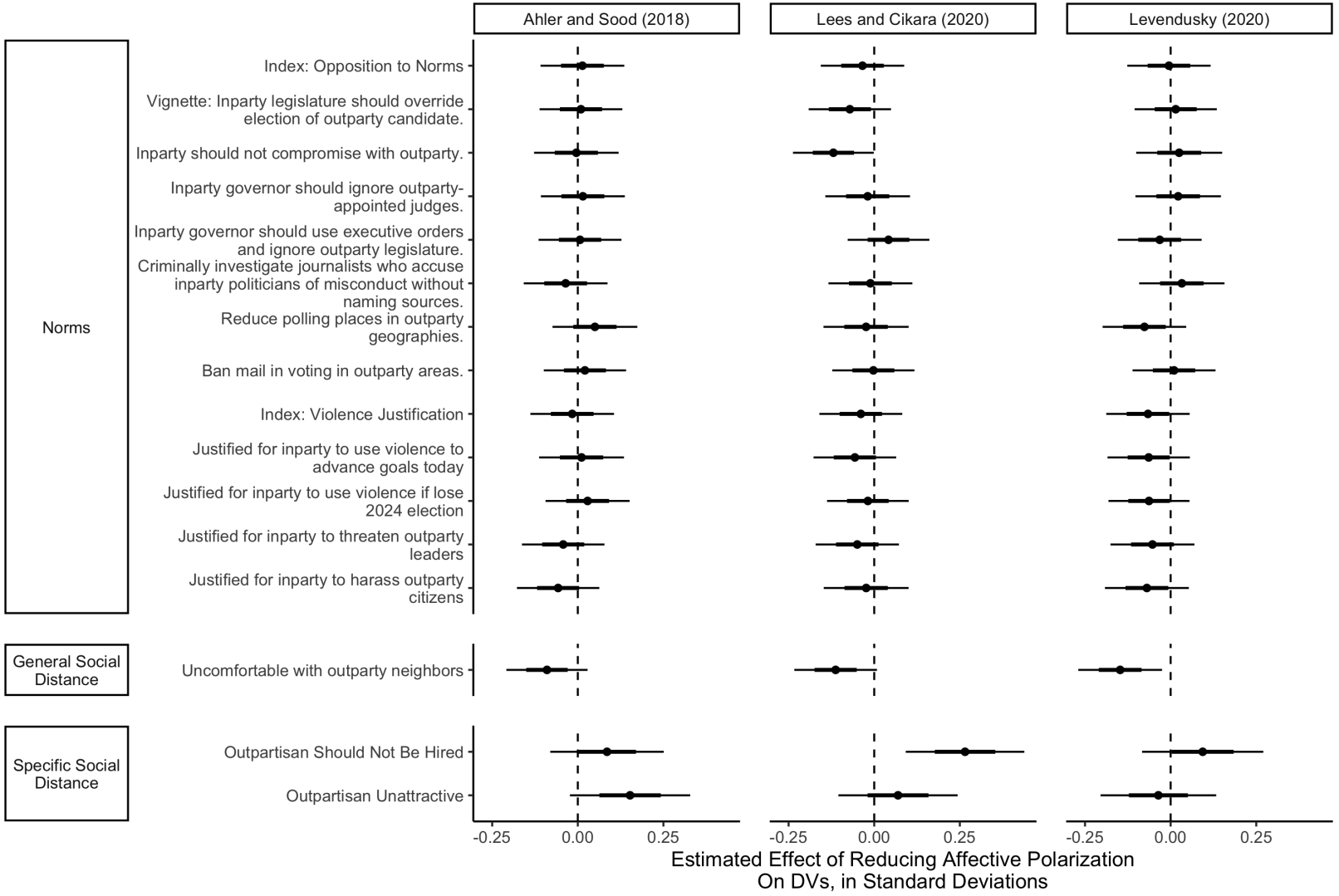
After asking the feeling thermometer questions necessary to measure affective polarization, we then asked two specific social distance items (respondents were randomly assigned to only one). Respondents were asked either to rate how likely they would be to hire a potential outpartisan job applicant (Iyengar and Westwood 2015) or to rate the attractiveness of an outpartisan dating profile (Nicholson et al. 2016) (see Figures B2 and B3 for full question wording).

As shown in Figure 11 and Figure B5, we find no effect of exogenously reducing affective polarization on these specific social distance items. If anything, reducing affective polarization may have increased partisan employment based discrimination, with treatment effects primarily driven by the Lees and Cikara (2020) manipulation.

[21] As we discuss further in Appendix D, and consistent with the results reported there, the negative trust game had no effect on affective polarization (effect on mass public affective polarization of 0.03 standard deviations; $p = 0.64$). We therefore do not include the negative trust game condition in Survey 5 in analyses examining effects on downstream outcomes.

We then asked respondents the norms index, election override vignette, and violence justification index, mirroring Survey 4. As shown in Figure 11 and Figure B5, we find no effect of exogenously reducing affective polarization on these downstream political attitudes. The results from Survey 5 suggest that the null effects we observe in Surveys 1-4 on political attitudes is unlikely to be explained by the artificial nature of the trust game experience. Three alternative interventions produce similar null findings.

Figure B5: Survey 5: Effects by Treatment



Notes: Point estimates are from multivariate regressions controlling for pre-registered covariates.

E.2 Testing for Demand Effects

Survey 5 also allows us to help assess whether experimenter demand effects may explain the null effects found in Surveys 1-4.

A first possible concern regarding demand is that the trust game did not actually manipulate affective polarization at all, but that respondents thought researchers wanted to see this occur, and so answered the survey questions as if it had manipulated affective polarization. We see two reasons to be doubtful that demand contaminated the original results. First, the downstream effects on general social distance items suggest that affective polarization itself was successfully manipulated. Second, as discussed in the paper and elaborated in Appendix D, the effects of the trust game are driven entirely by the positive treatment. If respondents simply "did what they thought researchers wanted," we would expect to see effects of both the negative and positive trust treatment. However, we only see effects of the positive treatment. With this said, Survey 5 helps address this further with its three new manipulations that reduce affective polarization. These three manipulations are less overt than the trust game manipulation and have also been demonstrated to effectively reduce affective polarization in other research. We continue to find consistent results with these three new manipulations as we did with the original trust game manipulation.

To examine this further, we also added a series of distractor questions to Survey 5. Adding space in the form of non-political distractor questions between the manipulations and our dependent variables may reduce respondent suspicion. We asked respondents how favorably they felt towards 6 brands: Crest, Colgate, Honda, Toyota, Ford, and Apple. Moreover, we randomly assigned whether these distractor questions came between the affective polarization manipulation and the feeling thermometer manipulation checks or not. We find that treatment effects of the manipulations on reducing affective polarization are indistinguishable whether or not affective polarization is measured before or after the distractor questions ($p = 0.98$).

A second possible concern is that respondents' affective polarization was manipulated, but that demand concealed its effects. For this to have led to null results that concealed the literature's predicted effects, a positive trust game experience would need to have led respondents to choose to misrepresent their attitudes in a manner consistent with less accountability and being more anti-democratic, cancelling out the salutary effects of the reduced affective polarization. This seems unlikely, but we examined this by asking respondents what they thought the purpose of the study was. In particular, at the very end of Survey 5, we asked respondents "Thank you for participating in this study. Lastly, we want to know what you think the purpose of this study was. Please provide your best guess in the box below." We sampled 1,000 responses to this question and manually read through them. Only three comments linked the manipulations to downstream attitudes, all three of which focused on interpersonal attitudes, not the downstream political measures, and all three of which were focused on the trust game. The three comments were: (1) "To see if your feelings about Republicans change when individual Republicans treat you well." (2) "To see if the amount of money we were given or not given relates to our reports of how we perceive others. I do not believe there was really another player." (3) "my reaction to democratic partners that were rather selfish." Overall, while some respondents may have seen through the manipulation, the rate was quite low. Moreover, we received no such comments among respondents assigned to the three new interventions in Survey 5; and none of the three comments we received referenced impacts on downstream political attitudes. This suggests it is unlikely that respondents en masse decided to present their attitudes in a more anti-democratic manner as a result of the manipulations that decreased their affective polarization (the direction of bias that would be needed for the effect of affective polarization to be cancelled out by demand effects).

F References for Online Appendix

References for Appendices

Ahler, Douglas J. and Gaurav Sood. 2018. "The parties in our heads: Misperceptions about party composition and their consequences." *Journal of Politics* 80(3):964–981.

Iyengar, Shanto and Sean J. Westwood. 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59(3):690–707.

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies* 76(3):1071–1102.

Lees, Jeffrey and Mina Cikara. 2020. "Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts." *Nature Human Behaviour* 4(3):279–286.

Levendusky, Matthew S. 2020. "Our common bonds: Using what Americans share to help bridge the partisan divide." *Unpublished Manuscript* .

Nicholson, Stephen P, Chelsea M Coe, Jason Emory and Anna V Song. 2016. "The politics of beauty: The effects of partisan bias on physical attractiveness." *Political Behavior* 38(4):883–898.