PNAS Nexus, 2024, 3, pgae286
https://doi.org/10.1093/pnasnexus/pgae286
Advance access publication 15 October 2024
Perspective

Explanations of and interventions against affective polarization cannot afford to ignore the power of ingroup norm perception

Zi Ting You [a] and Spike W. S. Lee [a,b,*]

[a]Department of Psychology, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada
[b]Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario M5S 3E6, Canada
*To whom correspondence should be addressed: Email: spike.lee@utoronto.ca
Edited By: Eugen Dimant

Abstract

Affective polarization, or animosity toward opposing political groups, is a fundamentally intergroup phenomenon. Yet, prevailing explanations of it and interventions against it have overlooked the power of ingroup norm perception. To illustrate this power, we begin with evidence from 3 studies which reveal that partisans' perception of their ingroup's norm of negative attitudes toward the outgroup is exaggerated and uniquely predicts their own polarization-related attitudes. Specifically, our original data show that in predicting affective polarization (i.e. how one feels about one's partisan outgroup), the variance explained by ingroup norm perception is 8.4 times the variance explained by outgroup meta-perception. Our reanalysis of existing data shows that in predicting support for partisan violence (i.e. how strongly one endorses and is willing to engage in partisan violence), ingroup norm perception explains 52% of the variance, whereas outgroup meta-perception explains 0%. Our pilot experiment shows that correcting ingroup norm perception can reduce affective polarization. We elucidate the theoretical underpinnings of the unique psychological power of ingroup norm perception and related ingroup processes. Building on these empirical and theoretical analyses, we propose approaches to designing and evaluating interventions that leverage ingroup norm perception to curb affective polarization. We specify critical boundary conditions that deserve prioritized attention in future intervention research. In sum, scientists and practitioners cannot afford to ignore the power of ingroup norm perception in explaining and curbing affective polarization.

Keywords: affect, polarization, intergroup relations, norm perception, meta-perception

Introduction

Our political climate today is fraught with toxic elements (1). One of the most potent toxins is affective polarization—partisans' animosity toward the opposing party's supporters (2). Such animosity exacerbates ideological polarization (3) and worsens behavioral dynamics across the aisle (4). Scholars have suggested that building connections between partisans might ameliorate affective polarization (2, 5). Unfortunately, in real life, affectively polarized partisans are not particularly motivated to build connections with each other. If anything, they are highly motivated to avoid interacting with those on the other side (6–10). With these social forces at work, how can researchers intervene?

We argue that affective polarization, as a partisan phenomenon, is fundamentally rooted in intergroup relations and social identity processes (11–14). To maximize the effectiveness of interventions, social and behavioral scientists must have a precise understanding of the most powerful drivers of affective polarization. That is, we must identify which aspects of intergroup relations and social identity exert the strongest influence. The goal of this Perspective piece is to point out that prevailing explanations of and interventions against affective polarization have missed the mark. Specifically, they have overlooked the power of ingroup norm perception.

Our perspective is motivated by evidence and theory. In the sections that follow, we begin with evidence from 3 studies (Section 1), all of which pit the effect of ingroup norm perception (i.e. perception of one's partisan ingroup members' thoughts and feelings) against that of outgroup meta-perception (i.e. perception of one's partisan outgroup members' thoughts and feelings). One study shows that in predicting affective polarization (i.e. how one feels about one's partisan outgroup), the variance explained by ingroup norm perception is 8.4 times the variance explained by outgroup meta-perception. Reanalysis of existing data shows that in predicting support for partisan violence (SPV) (i.e. how strongly one endorses and is willing to engage in partisan violence), ingroup norm perception explains 52% of the variance, whereas outgroup meta-perception explains 0%. A pilot experiment shows that correcting ingroup norm perception reduces Republicans' affective polarization, whereas correcting outgroup meta-perception does not.

Competing Interest: The authors declare no competing interests.
Received: April 3, 2024. **Accepted:** June 30, 2024

To make sense of such evidence, we elucidate the theoretical underpinnings of the unique psychological power of ingroup norm perception and related ingroup processes (Section 2). Building on these empirical and theoretical analyses, we propose approaches to designing and evaluating interventions that leverage ingroup norm perception to curb affective polarization (Section 3). Our proposed approaches are inspired by various effective interventions that have leveraged intergroup processes (but not ingroup norm perception) to curb affective polarization or those that have leveraged ingroup norm perception to change diverse attitudinal and behavioral outcomes (but not affective polarization). We specify critical boundary conditions, derived from theoretical principles, that deserve prioritized attention in future intervention research.

Ingroup norm perception trumps outgroup meta-perception in driving affective polarization

Affective polarization is a phenomenon that involves group-level psychological processes, not just individual-level ones. Partisans dislike members of the opposing party often because of their party affiliation more than because of their individual characteristics (15). Highlighting the "groupiness" of partisan dynamics, we bring insights from intergroup relations and social identity theory (11) to bear on our analysis of affective polarization. From the intergroup literature, what processes exert robust psychological and behavioral influence?

Classic intergroup research (16) has found that simple group categorization—seeing other individuals as either "us" (ingroup) or "them" (outgroup)—is sufficient to cement bias toward the outgroup. This characterizes the reality of current political group relations (13, 15). Partisans value being part of a group of politically similar others. They define part of their self-identity by their membership in and belongingness to their political ingroup. Those outside this group, or politically dissimilar others, are outgroup members and targets of bias.

In the bipartisan context of the United States, one's political ingroup vs. outgroup typically comprises those who support the same vs. opposing political party. With only 2 major parties, group categorization is particularly straightforward, salient, and easy. For instance, partisans can discern even from still photos whether a target is more likely a Democrat or Republican (17, 18). Easy group categorization, unfortunately, comes with difficult intergroup relations. With only 2 major parties, political group relations are prone to being construed as a zero-sum competition because "their win equals our loss" (19), resulting in hypersensitivity to the relative successes of one's political ingroup vs. outgroup.

These dynamics of American politics bring group membership to the forefront. When ingroup identities and fellow ingroup members are highly valued (20), psychological processes related to the ingroup exert powerful influence on how people think, feel, and act toward outgroup members (21). For example, individuals tend to assimilate to what they perceive to be their ingroup members' normative attitudes and behaviors. Throughout this paper, by "normative," we mean descriptive norms ("what others do"; as opposed to "what should be done," which would be injunctive norms) (22, 23). We contend that these processes are especially relevant to political group dynamics.

Given the well-established impact of ingroup processes, one might expect their role to be prominently featured in psychological research on affective polarization. That is not the case. Psychological research on affective polarization has largely ignored the power of individuals' perception of their ingroup's attitudes toward the outgroup, despite how strongly the ingroup figures into individuals' social identity. To be clear, existing explanations of affective polarization have drawn on intergroup processes in general, but have not incorporated how individuals adopt their ingroup members' attitudes in particular.

Consider a recent popular paradigm in this area of research: outgroup meta-perception. It focuses on one's perception of the partisan outgroup's attitudes toward one's partisan ingroup (24, 25). Outgroup meta-perception has several properties. First, it tends to be negatively exaggerated, i.e. people tend to have an exaggerated perception of how negatively their outgroup feels toward their ingroup. Second, negatively exaggerated outgroup meta-perception tends to worsen people's feelings toward the outgroup. Third, presenting accurate information to correct exaggerated outgroup meta-perception tends to improve feelings toward the outgroup.

While these properties are valuable and have inspired our own thinking, outgroup meta-perception is fundamentally different from ingroup norm perception. The two kinds of perception imply different roots of affective polarization and suggest different foci of intervention. Explaining affective polarization in terms of outgroup meta-perception implies that affective polarization is particularly sensitive to intergroup dynamics and is reactive to one's perception of the outgroup's attitudes (26). Explaining affective polarization in terms of ingroup norm perception implies that one's negativity toward the outgroup can emerge from dynamics within one's ingroup itself, without being a response to one's perception of the outgroup's negativity toward one's ingroup.

Juxtaposing ingroup norm perception and outgroup meta-perception raises a simple empirical question: which one matters more for affective polarization? To find out, we first analyze data from 2 studies that measure both ingroup norm perception and outgroup meta-perception within each participant and thus allow us to compare their unique predictive effects on polarization-related outcomes.

Our original data

We collected and analyzed original data from U.S. partisans (123 Democrats, 114 Republicans, $N = 237$, attaining 87.46% statistical power in detecting an effect size of Cohen's $f^2 = 0.05$ at $\alpha = 0.05$ with 2 predictors in the regression model) (see Supplementary Materials and Methods for sample characteristics). Using the feelings thermometer, participants rated how positive and warm they felt toward various social groups, including Democrats and Republicans (among others), from 0 = very cold to 100 = very warm. Participants also rated how they perceived the average Democrat and the average Republican to feel toward the same social groups, using the same feelings thermometer.

Altogether these ratings allowed us to assess 3 constructs of focal interest on the same metric: (i) each participant's own feeling toward their partisan outgroup, (ii) their ingroup norm perception (i.e. perception of their partisan ingroup's feeling toward their partisan outgroup), and (iii) their outgroup meta-perception (i.e. perception of their partisan outgroup's feeling toward their partisan ingroup). Descriptive statistics are provided in Table S1 and zero-order correlations in Table S2.

Did ingroup norm perception and outgroup meta-perception predict participants' actual feelings toward their partisan outgroup? If so, which one was a stronger predictor? We mean-centered the variables and regressed actual feelings toward one's partisan outgroup on both ingroup norm perception and outgroup meta-perception (Table 1, model 1), then added control

predictors (model 2) including partisan affiliation, its interaction with ingroup norm perception, and its interaction with outgroup meta-perception.

In model 1, actual feeling toward one's partisan outgroup was significantly predicted by ingroup norm perception ($\beta = 0.45$, $P < 0.001$; all regression coefficients reported in text are standardized) and outgroup meta-perception ($\beta = 0.16$, $P = 0.021$). However, the coefficient for ingroup norm perception was significantly larger than the coefficient for outgroup meta-perception, $F_{1,\ 234} = 6.07$, $P = 0.015$. The variance explained by ingroup norm perception (13.24%) was 8.4 times the variance explained by outgroup meta-perception (1.58%).

After adding control predictors (model 2), actual feeling toward one's partisan outgroup remained predicted by ingroup norm perception ($\beta = 0.45$, $P < 0.001$) and outgroup meta-perception ($\beta = 0.15$, $P = 0.027$) in the same way as in model 1. These effects were comparable (i.e. did not differ significantly) between Democrat and Republican participants (ingroup norm perception × partisan affiliation, $\beta = 0.01$, $P = 0.858$; outgroup meta-perception × partisan affiliation, $\beta = -0.03$ $P = 0.609$). There was a mean-level partisan difference such that Republican participants felt warmer toward Democrats than vice versa (main effect of partisan affiliation, $\beta = 0.14$, $P = 0.010$). But as far as our primary interest in the predictive effects of ingroup norm perception and outgroup meta-perception is concerned, actual feeling toward one's partisan outgroup was predicted most strongly by ingroup norm perception among both Democrats and Republicans.

To check for robustness, we conducted the same regression analyses using a more complex operationalization of the 3 constructs—as difference scores—in accordance with existing research on affective polarization (27). Specifically, we operationalized (i) out-partisan feeling as one's own feeling toward one's partisan outgroup minus one's own feeling toward one's partisan ingroup, (ii) ingroup norm perception as perception of one's partisan ingroup's feeling toward one's partisan outgroup minus perception of one's partisan ingroup's feeling toward one's partisan ingroup, and (iii) outgroup meta-perception as perception of one's partisan outgroup's feeling toward one's partisan ingroup minus perception of one's partisan outgroup's feeling toward one's partisan outgroup. Descriptive statistics are provided in Table S1, zero-order correlations in Table S2, and regression results in Table S3. Actual differential feeling toward one's partisan outgroup (over one's partisan's ingroup) was significantly predicted by ingroup norm perception ($\beta = 0.30$, $P < 0.001$ without control predictors; $\beta = 0.31$, $P < 0.001$ with control predictors), not by outgroup meta-perception ($\beta = 0.09$, $P = 0.222$ without control predictors; $\beta = 0.08$, $P = 0.321$ with control predictors).

Overall, our original data showed that ingroup norm perception had strong and robust predictive effects on actual feelings toward partisan outgroup members, whereas outgroup meta-perception had weaker and less robust predictive effects. These patterns emerged among both Democrats and Republicans. Collinearity diagnostics (Table S4) found no concern of multicollinearity (28, 29). As a reality check, additional analyses (Supplementary Results) confirmed that both Democrats and Republicans showed affective polarization (i.e. feeling cold toward the partisan outgroup), negatively exaggerated ingroup norm perception (i.e. perceiving their partisan ingroup's feelings toward the partisan outgroup to be worse than reality), and negatively exaggerated outgroup meta-perception (i.e. perceiving their partisan outgroup's feelings toward the partisan ingroup to be worse than reality). Even though both ingroup norm perception and outgroup meta-perception were negatively exaggerated, when pitted against each other, it was ingroup norm perception that showed much stronger predictive effects.

Our reanalysis of existing data

To assess the replicability of our findings, we looked for additional data from existing studies that would allow us to pit the predictive effects of ingroup norm perception and outgroup meta-perception against each other in a within-participant design. We found one study: a 2022 *PNAS* paper by Mernyk et al. (30), whose study 1 met our criterion, with publicly available data (31).

The original authors focused on examining the effect of outgroup meta-perception on SPV. They hypothesized that participants' own SPV toward their political outgroup was a response to exaggerated meta-perception of their political outgroup's SPV toward their political ingroup. The original authors also measured ingroup norm perception of SPV but did not compare its effect against that of outgroup meta-perception on participants' own SPV. We reanalyzed their data to test this comparison.

Before presenting the results, we should note that SPV toward one's partisan outgroup is more severe than just feeling cold toward them. It is plausible that the two might show different patterns of results. For example, even if a Democrat felt cold toward Republicans, they might, due to social desirability concerns, be uncomfortable explicitly endorsing SPV statements such as "it is justified for Democrats to use violence in advancing their political goals these days" and "[it is] OK for an ordinary Democrat in the public to harass an ordinary Republican on the internet, in a way that makes the target feel frightened." Our reanalysis of Mernyk et al.'s (30) data explored whether SPV toward one's partisan outgroup was predicted by ingroup norm perception and outgroup meta-perception in the same way that cold feelings toward one's partisan outgroup was predicted by ingroup norm perception and outgroup meta-perception in our original data.

We applied the same filters and attention checks as in the original authors' analyses, resulting in 702 participants (354 Democrats, 348 Republicans) (see Supplementary Materials and Methods for sample characteristics). The sample size would attain 99.98% statistical power in detecting an effect size of Cohen's $f^2 = 0.05$ at $\alpha = 0.05$ with 2 predictors in the regression model. Descriptive statistics of the 3 constructs of focal interest are provided in Table S5 and zero-order correlations in Table S6.

Similar to our original study, reality checks (Supplementary Results) confirmed that both ingroup norm perception of SPV and outgroup meta-perception of SPV were exaggerated (i.e. perceptions were graver than reality). Ingroup norm perception was less exaggerated, or more accurate, than outgroup meta-perception. This particular point led the original authors to offer a suggestion that we consider understandable but a missed opportunity. They suggested that the higher accuracy of ingroup norm perception "leaves little room for a correction intervention" (30). Perhaps because of that, they did not explore the predictive effect of ingroup norm perception any further, although the analytic code provided in their online Supplementary Material included a t test showing that ingroup norm perception of SPV was stronger than actual SPV. We reran their code and reproduced what they found (Supplementary Results). It suggested that ingroup norm perception, while more accurate than outgroup meta-perception, still exaggerated partisans' actual SPV.

Recognizing the slightly exaggerated ingroup norm perception and more exaggerated outgroup meta-perception of SPV, did both kinds of perception predict participants' actual SPV toward their partisan outgroup? We mean-centered the

Table 1. Linear multiple regression models for our original data.

Predictor	Model 1				Model 2			
	β (SE)	95% CI	B (SE)	P	β (SE)	95% CI	B (SE)	P
Intercept	0.00 (0.05)	−0.11 to 011	29.73 (1.50)	<0.001	0.00 (0.05)	−0.10 to 011	29.93 (1.49)	<0.001
Ingroup norm perception	**0.45**[a] (0.07)	**0.32 to 058**[a]	**0.66**[a] (0.10)	**<0.001**[a]	**0.45** (0.07)	**0.32 to 058**[a]	**0.66**[a] (0.10)	**<0.001**[a]
Outgroup meta-perception	**0.16**[a] (0.07)	**0.02 to 029**[a]	**0.23**[a] (0.10)	**0.021**[a]	**0.15**[a] (0.07)	**0.02 to 028**[a]	**0.22**[a] (0.10)	**0.027**[a]
Partisanship (−1 = Democrat, 1 = Republican)					**0.14**[a] (0.05)	**0.03 to 025**[a]	**3.85**[a] (1.49)	**0.010**[a]
Ingroup norm perception × Partisanship					0.01 (0.07)	−0.12 to 014	0.02 (0.10)	0.858
Outgroup meta-perception × Partisanship					−0.03 (0.7)	−0.17 to 010	−0.05 (0.10)	0.609
Observations | R^2/Adjusted R^2	237 | 0.312/0.306				237 | 0.332/0.318			

Cross-partisan feeling was regressed on ingroup norm perception and outgroup meta-perception (model 1), together with control predictors (model 2). Each kind of perception was centered around its sample mean. Model comparison found that models 1 and 2 did not differ significantly in variance explained, $\chi^2_3 = 2.33$, $P = 0.075$.
Abbreviation: CI = confidence interval.
[a]Bold values indicate significant predictors (other than intercept).

Table 2. Linear multiple regression models for our reanalysis of Mernyk et al. (30), study 1.

Predictor	Model 3				Model 4			
	β (SE)	95% CI	B (SE)	P	β (SE)	95% CI	B (SE)	P
Intercept	0.00 (0.03)	−0.05 to 005	9.79 (0.50)	<0.001	0.00 (0.03)	−0.05 to 005	9.81 (0.50)	<0.001
Ingroup norm perception	**0.74**[a] (0.03)	**0.69 to 079**[a]	**0.70**[a] (0.02)	**<0.001**[a]	**0.74**[a] (0.03)	**0.69 to 079**[a]	**0.70**[a] (0.03)	**<0.001**[a]
Outgroup meta-perception	0.00 (0.03)	−0.05 to 005	0.00 (0.02)	0.947	0.01 (0.03)	−0.05 to 006	0.00 (0.02)	0.857
Partisanship (−1 = Democrat, 1 = Republican)					−0.01 (0.03)	−0.06 to 004	−0.13 (0.50)	0.799
Ingroup norm perception × Partisanship					0.00 (0.03)	−0.05 to 005	0.00 (0.03)	0.986
Outgroup meta-perception × Partisanship					−0.03 (0.03)	−0.09 to 002	−0.02 (0.02)	0.197
Observations | R^2/Adjusted R^2	702 | 0.548/0.547				702 | 0.549/0.546			

Support for political violence toward political outgroup was regressed on ingroup norm perception and outgroup meta-perception (model 3), together with control predictors (model 4). Each kind of perception was centered around its sample mean. Model comparison found that models 3 and 4 did not differ significantly in variance explained, $\chi^2_3 = 0.61$, $P = 0.608$.
Abbreviation: CI = confidence interval.
[a]Bold values indicate significant predictors (other than intercept).

variables and regressed actual SPV toward one's partisan outgroup on both ingroup norm perception and outgroup meta-perception (Table 2, model 3), then added control predictors (model 4) including partisan affiliation, its interaction with ingroup norm perception, and its interaction with outgroup meta-perception.

In model 3, actual SPV toward one's partisan outgroup was significantly predicted by ingroup norm perception ($\beta = 0.74$, $P < 0.001$) but not by outgroup meta-perception ($\beta = 0.00$, $P = 0.947$). The coefficient for ingroup norm perception was significantly larger than the coefficient for outgroup meta-perception, $F_{1, 699} = 330.84$, $P < 0.001$. The variance explained by ingroup norm perception was 52.22%; the variance explained by outgroup meta-perception was 0%.

The strong predictive effect of ingroup norm perception was robust to the addition of control predictors, as shown in model 4, in which actual SPV toward one's partisan outgroup remained significantly predicted by ingroup norm perception ($\beta = 0.74$, $P < 0.001$) but not by outgroup meta-perception ($\beta = 0.01$, $P = 0.857$). These effects were comparable (i.e. did not differ significantly) between Democrat and Republican participants (ingroup norm perception × partisan affiliation, $\beta = 0.00$, $P = 0.986$; outgroup meta-perception × partisan affiliation, $\beta = -0.03$, $P = 0.197$). Mean levels of actual SPV toward one's partisan outgroup were comparable between Democrat and Republican participants (main effect of partisan affiliation, $\beta = -0.01$, $P = 0.799$). Collinearity diagnostics (Table S4) again found no concern of multicollinearity (28, 29). These results suggest that among both Democrats and Republicans, actual SPV toward one's partisan outgroup was similar in level and similarly predicted by ingroup norm perception, not by outgroup meta-perception.

The cross-sectional and correlational nature of the two studies analyzed here provide convergent evidence for the strong predictive effect of ingroup norm perception. But neither study manipulated it to test its causal effect. It is possible that partisans adjust their perceived ingroup norm to match their own attitudes (instead of basing their own attitudes on their perceived ingroup norm), a process that Mernyk et al. (30) also suggested as a possible explanation for the correlation between partisans' ingroup norm perception of SPV and their own SPV. To provide causal evidence, we conducted a pilot experiment.

Our pilot experiment

To examine the causal effects of ingroup norm perception and outgroup meta-perception, we conducted a pilot experiment using a 2 (correcting vs. not correcting ingroup norm perception) × 2 (correcting vs. not correcting outgroup meta-perception) between-participant design. We collected and analyzed data from U.S. partisans (216 Democrats, 216 Republicans, $N = 432$) (see Supplementary Materials and Methods for sample characteristics).

Using the feelings thermometer, participants rated how positive and warm they perceived the average Democrat and the average Republican to feel toward various social groups, including Democrats and Republicans (ingroup norm perception and outgroup meta-perception) among other groups. Next, depending on the condition, participants were presented with either correct

ingroup norm perception data from the original study described previously, or correct outgroup meta-perception data from the same study, or both, or neither. After the manipulation, participants rated how they felt toward these same groups (actual feelings) among other variables. In accordance with existing research on affective polarization (27) and our previous study (Section 1a), we operationalized out-partisan feelings, ingroup norm perception, and outgroup meta-perception as difference scores. Descriptive statistics are provided in Table S7.

Correcting ingroup norm perception improved Republicans' actual feeling toward their partisan outgroup (estimate = 20.52, SE = 7.92, $t_{424} = 2.59$, $P = 0.010$). This effect remained significant even after controlling for premanipulation baseline levels of ingroup norm perception and outgroup meta-perception (estimate = 17.86, $SE = 6.76$, $t_{416} = 2.64$, $P = 0.009$). Correcting outgroup meta-perception did not improve Republicans' actual feeling toward their partisan outgroup (estimate = 1.73, $SE = 7.92$, $t_{424} = 0.22$, $P = 0.828$).

Like Republicans, Democrats' actual feeling toward their partisan outgroup was not significantly affected by correcting outgroup meta-perception (estimate = −1.12, $SE = 8.01$, $t_{424} = -0.14$, $P = 0.889$). But unlike Republicans, Democrats' actual feeling toward their partisan outgroup was not significantly affected by correcting ingroup norm perception (estimate = −9.96, $SE = 8.01$, $t_{424} = -1.24$, $P = 0.214$). This discrepancy between Republicans and Democrats might have arisen because the corrective information used in the ingroup norm perception manipulation was stronger for Republicans than for Democrats: Republicans read that "actual Republicans felt between 36% to 51% warmer [toward Democrats] than the average Republican's guess," whereas Democrats read that "actual Democrats felt between 18% and 24% warmer [toward Republicans] than the average Democrat's guess." This illustrates one of the challenges entailed by operationalizing the manipulation based on real data from prior studies, a theme we revisit in Section 3c.

Conceptually replicating the predictive effects in our original data, across Republicans and Democrats, actual feeling toward one's partisan outgroup was significantly predicted by baseline levels of ingroup norm perception ($\beta = 0.49$, $SE = 0.044$, $t_{416} = 11.28$, $P < 0.001$) and outgroup meta-perception ($\beta = 0.090$, $SE = 0.043$, $t_{416} = 2.11$, $P = 0.036$). Again, the coefficient for ingroup norm perception was significantly larger than the coefficient for outgroup meta-perception, $F_{1, 416} = 31.61$, $P < 0.001$. The variance explained by ingroup norm perception (19.93%) was 28.5 times the variance explained by outgroup meta-perception (0.70%).

In short, correcting ingroup norm perception improved Republicans' actual feeling toward their partisan outgroup. Correcting outgroup meta-perception did not. Pitting baseline levels of ingroup norm perception and outgroup meta-perception against each other, both Republicans' and Democrats' actual feeling toward their partisan outgroup was much more strongly predicted by ingroup norm perception than by outgroup meta-perception.

Summary of our original data, reanalysis of existing data, and pilot experiment

To our knowledge, no existing work has specifically examined ingroup norm perception as a driver of affective polarization, let alone pitted it against outgroup meta-perception. Our original data and our reanalysis of existing data converge in providing the first evidence that ingroup norm perception has unique, large, and robust predictive effects on partisan animosity and support for partisan violence.

Note that in both studies, ingroup norm perception is descriptively *less* exaggerated than outgroup meta-perception (Supplementary Results), echoing prior observations—experiment 4 and supplemental experiment A in Lees and Cikara (26), experiment 4 in Ruggeri et al. (32)—and prior suggestions that the two kinds of perception might be subject to different types of inaccuracy (24). The finding that ingroup norm perception is descriptively less exaggerated than outgroup meta-perception might lead to the tempting but erroneous conclusion that ingroup norm perception matters less than outgroup meta-perception. The opposite is true. Ingroup norm perception matters more, at least in terms of predicting polarization-related outcomes. The less exaggerated nature of ingroup norm perception might even be reinterpreted as suggesting that individuals' feelings toward their partisan outgroup track what they perceive their ingroup members to feel more closely than what they perceive their outgroup members to feel.

Our findings represent an extension and indirect challenge to some nuanced evidence from related work (26), which found that a condition that corrected both ingroup norm perception and outgroup meta-perception changed participants' own perception of obstructionism to the same extent as a condition that corrected only outgroup meta-perception. In other words, adding correction of ingroup norm perception to correction of outgroup meta-perception did not exert any additional effect. This study, however, did not include any condition that corrected only ingroup norm perception (without correcting outgroup meta-perception), rendering direct comparisons difficult. Our pilot experiment teased apart the causal effects of correcting ingroup norm perception and correcting outgroup meta-perception. We found that correcting ingroup norm perception improved Republicans' actual feeling toward their partisan outgroup. Correcting outgroup meta-perception did not.

Why does ingroup norm perception have such powerful effects? Theoretical underpinnings of its unique explanatory power deserve unpacking. We do so in the next section by contextualizing ingroup norm perception in robust psychological forces of intergroup relations and social identity.

Theoretical underpinnings of the psychological power of ingroup norm perception and related ingroup processes

Theorizing by political psychologists has danced around the notion of ingroup norm perception but not directly recognized its importance in driving affective polarization. We first review why ingroup norms should exert powerful influence on group members' attitudes toward the outgroup, invoking mechanisms of ingroup conformity and opinion polarization. Afterward, to fully appreciate the power of ingroup norms, we provide a critical appraisal of existing explanations of affective polarization through the lens of normative processes.

Why ingroup norm perception should drive affective polarization

We argue that affective polarization is driven and exacerbated by partisans mirroring what they perceive to be their ingroup's norm of disliking the outgroup. A key tenet of intergroup psychology is that individuals tend to conform to fellow ingroup members. This is because the ingroup is highly valued and seen as part of

the self (16). Individuals seek to affirm their ingroup identity and assimilate. They follow what they perceive to be their ingroup's norms in adopting and expressing attitudes (21). Indeed, causal effects of ingroup norms on personal attitudes have been documented in numerous domains (33).

For example, expressing and suppressing prejudice toward social groups are products of conforming to perceived group norms. The more people perceive that it is socially normative and acceptable to hold negative feelings toward a given social group, the more negative they themselves report feeling toward that group (34). These processes have been found to underlie regional variations in prejudice. The more people perceive others in their region to condone prejudice, the more people express prejudicial attitudes themselves (35).

Applying these principles to the political domain, if partisans perceive a norm of fellow co-partisans feeling negatively toward the opposing party, they are inclined to adjust their own attitudes to match that norm. Unfortunately, partisans' perception of the political ingroup's normative attitudes toward the outgroup tends to be negatively exaggerated, as illustrated in our analyses. Such exaggeration can be attributed to 2 cognitive processes.

First, individuals tend to perceive their group's average attitudes as being more extreme than their own attitudes (36, 37). This tendency emerges because group members are motivated to distance themselves from the outgroup to maintain the value and distinctiveness of the ingroup. They thus exaggerate differences between the ingroup and the outgroup, leading to a more extreme ingroup norm perception (36). Moreover, group members often base their ingroup norm perception on the prototypical ingroup member, such as a politician with frequent media coverage, whose attitudes might be more extreme than the average group member (27, 37).

Second, partisans are especially negatively biased when estimating their feelings and attitudes toward their partisan outgroup. They predict experiencing more negative affect upon encountering the outgroup's opinions than they actually report experiencing when it happens (38). They also estimate that they will feel more negative if an outgroup politician succeeds in an upcoming election than they actually report feeling after the election (39, 40).

These processes can contribute to partisans' negatively exaggerated perception of the political ingroup's normative attitudes toward the outgroup. Insofar as partisans match their own attitudes toward the outgroup to their perceived ingroup norm, they will end up harboring more negative attitudes than the average group member. If each group member does the same, the average ingroup attitude itself will drift negative, intensifying affective polarization (37). Making matters worse, such group dynamics could form a feedback loop. More affectively polarized individuals tend to follow the ingroup's norms more (41). Following the norm further polarizes group members, through the processes discussed previously, creating more polarized group members who now follow an even more extreme perceived norm, further exacerbating affective polarization.

These ingroup norm processes should not be foreign to political psychology, but they have not been comprehensively integrated into theories and research on affective polarization, despite their demonstrated role in a variety of politically related outcomes. For example, research on partisan cue receptivity has found that cues of fellow political ingroup members' issue positions constitute an important influence on one's own political opinions (42, 43). In laboratory contexts, cues of ingroup norms can flip conservatives' and liberals' opposition to the same politically neutral policy, simply by telling participants that either conservative or liberal others already supported it (44). In naturalistic contexts, the ingroup's influence on political attitudes persists even for individuals who have prior knowledge about a policy (45). Crucial political behaviors, such as voting, are also influenced by perceived ingroup norms (46, 47).

Ingroup norm processes, well studied as they are in relation to various political attitudes and behaviors, remain underappreciated in the realm of affective polarization. Research in this area, particularly over the past several years, has focused instead on outgroup meta-perception. Although ingroup norm perception may appear to resemble outgroup meta-perception in structure, it is important to realize that ingroup-focused and outgroup-focused psychological processes often feed into different group-based biases. For example, favoritism of partisan ingroup and derogation of partisan outgroup are separate psychological processes that do not necessarily co-occur (20, 48). When given the choice, partisans prefer to help their ingroup members rather than to harm their outgroup members, illustrating their attachment to and prioritization of the political ingroup (49, 50).

Both ingroup norm perception and outgroup meta-perception reflect perception of other individuals' attitudes, but outgroup meta-perception does not capture and certainly is not reducible to ingroup norm perception. If it did and if it were, our analyses in Section 1 would have shown that outgroup meta-perception trumps ingroup norm perception in driving polarization-related outcomes. But we found the opposite. Ingroup norm perception trumps outgroup meta-perception in explaining far more variance of partisan animosity and support for partisan violence.

The unique and powerful impact of ingroup norm perception on affective polarization, we submit, is rooted in partisans' valued ingroup identity and, as a corollary, their motivation to conform and assimilate to their perception of their ingroup members' attitudes. Unfortunately, when it comes to estimating one's political ingroup's negative attitudes toward the outgroup, perception tends to exaggerate reality. Exaggerated ingroup norm perception could beget more negative feelings, which could beget more negative ingroup norm perception, resulting in a vicious cycle.

Affective polarization research has recognized the importance of other ingroup processes but not ingroup norm perception

Although there is a lack of research applying ingroup norm perception to affective polarization, existing explanations of affective polarization do draw on other ingroup processes. The field evidently recognizes the importance of the political ingroup in general but not ingroup norm perception in particular. To appreciate its central role, we offer a critical appraisal of several existing explanations of affective polarization through the lens of ingroup norm processes.

One account for the rise in affective polarization is that the boundaries between political groups have become clearer in recent times, because other identities have aligned more clearly with partisan identities. This phenomenon, called ideological sorting, makes it easier to identify ingroup and outgroup members (51). Researchers most commonly study sorting in the context of Democrats being more likely to lean liberal and Republicans conservative (48, 52, 53), but religious and racial identities have also become more strongly sorted along party lines (54, 55). The concomitant rise in ideological sorting and rise in affective polarization have prompted speculation of a relation between the two (51).

We argue that ideological sorting itself could result from ingroup norm processes. There are signs that partisans shift their ideological opinions to align with their fellow ingroup members'. As already noted, many voters' ideological opinions are not driven

by cohesive worldviews (43). Rather, naïve voters often take cues from fellow ingroup members (56). Indeed, cross-lagged panel models have found that ideological sorting is reciprocally associated with affective polarization, suggesting that ideological cohesion in political groups predicts more outgroup bias, and more outgroup bias in turn predicts more cohesive groups (51). These findings are compatible with the possibility that sorting heightens affective polarization through partisans' attachment to their political ingroup.

Mass media or social media filter bubbles are another catalyst for affective polarization that can act through ingroup norm perception (57–59). Watching partisan media from an ingroup-leaning source has been found to heighten affective polarization through 2 compounding mechanisms (59). First, the media messaging and content intensify negative emotions toward the outgroup. Second, these sources use ingroup members to deliver the message, increasing the subjective value and trustworthiness of the news content (60).

Complementing these recognized mechanisms, we add that partisan media can also deliver unambiguous information about ingroup norms. The medium itself cues ingroup norms, as the actual or perceived partisan leaning of each medium reflects its group affiliation. Viewers are sensitive to these cues (61). When partisan media report on political polarization, viewers consequently believe the electorate is more polarized and they themselves dislike the outgroup more (62). In other words, partisan media instigate the process of shifting individual attitudes to match perceived ingroup norms.

Partisans also avoid cross-party media sources and interactions with the outgroup. A lack of interactions with outgroup members is typically associated with more negative feelings toward and negative stereotypes of the outgroup, so partisans' avoidance contributes to affective polarization (63, 64). Unfortunately, partisans' interest in hearing about the outgroup's opinions is about as high as their interest in taking out the trash; partisans will even give up a monetary reward to avoid outgroup information (8). Online, they choose to avoid content that cuts across party lines (65). It has been suggested that partisans avoid these cross-cutting opportunities because they dislike the outgroup or political discussions in general (66).

Another reason for such avoidance, we argue, is that partisans perceive it to be normative. Group members conform to perceived norms of their ingroup's attitudes and behaviors. When it comes to intergroup contact, those who believe contact is normative intend to have more contact (67); equivalently, those who believe that contact is non-normative intend to have less contact. Positive intergroup interactions are more common in regions where people perceive positive intergroup contact to be common (35). Intergroup contact and outgroup bias may exhibit a reciprocal relationship, such that partisans who are less polarized are more amenable to cross-party interactions. Regardless of the precise nature of this relationship, the evidence is clear that normative influences are strong predictors of contact intentions (68).

The prevailing explanations of affective polarization surveyed previously already draw on various ingroup processes that emerge from political identity as a social identity group. We argue that these explanations can be construed through the lens of ingroup norm perception. To be clear, we are not arguing that all these explanations have equal epistemic status or drive affective polarization to the same extent. For example, the extent to which selective media exposure influences affective polarization remains debated (2). We do argue that there are strong theoretical bases underlying the contention that ingroup norm perception is a key driver of affective polarization. By implication, if we can manipulate ingroup norm perception, it would be a potent breakpoint for intervention. We pursue this direction in the next section.

Leveraging ingroup norm perception for interventions against affective polarization

Building on our empirical and theoretical analyses that have highlighted the unique psychological power of ingroup norm perception, we propose approaches to designing and evaluating interventions that leverage it to curb affective polarization. Our proposed approaches are inspired by various interventions that have successfully leveraged intergroup processes (but not ingroup norm perception) to curb affective polarization and interventions that have leveraged ingroup norm perception to produce effective changes in diverse attitudinal and behavioral outcomes (but not affective polarization). We first review these effective interventions, then describe our proposed approaches, and finally identify critical boundary conditions, derived from theoretical principles, that deserve prioritized attention in future intervention research.

Some effective interventions have leveraged intergroup processes (but not ingroup norm perception) to curb affective polarization

Many existing interventions against affective polarization already leverage intergroup psychological phenomena. They show that intergroup processes are often more effective than directly targeting stereotypes or motivations to express less prejudice. The most successful interventions tend to model established interventions against bias in other group domains.

For example, reminding partisans of their shared identity as Americans reduces affective polarization by way of evoking a common ingroup identity (69, 70). Facilitating intergroup contact (63) or positive social interactions and intimate friendships between members of opposing political groups yields effect sizes typical of contact in other domains (71–74). Even reading about other ingroup members having a positive interaction (i.e. "extended intergroup contact") reduces affective polarization (75). Finally, outgroup meta-perception interventions, by targeting partisans' reactive dislike of the outgroup, reduce perceptions of obstructionism (26, 32), dehumanization (76), animus (77), and support for partisan violence (30) toward the opposing party.

These examples demonstrate the utility of leveraging intergroup processes to reduce affective polarization in different forms. Meanwhile, we find the absence of ingroup norm perception interventions from this body of work glaring. It is also surprising because individuals' norm perceptions are not set in stone. They are conducive to intervention, with promising results already shown in other domains (78), as described next.

Other effective interventions have leveraged ingroup norm perception to change diverse attitudinal and behavioral outcomes (but not affective polarization)

While yet to be tested on affective polarization, variants of norm perception interventions, sometimes called social comparison interventions (79), have effectively changed sticky attitudes and consequential behaviors. A classic study showed that, over the course of a college semester, male undergraduate students changed their attitudes toward alcohol consumption on campus to match their perception of their peers' attitudes by the end of the semester (80). Follow-up studies corrected perception of the

norm and successfully reduced college students' alcohol consumption (81, 82). Another successful norm-based intervention found that Californian residents used less energy, as measured by objective household electricity meter readings, if they had received a doorhanger message saying that their neighbors had engaged in energy-conserving behaviors than if they had received doorhanger messages describing the personal and social benefits of energy conservation (83, 84). Even vehement beliefs, such as belief in vaccination conspiracies during COVID-19, were susceptible to norm-based messaging interventions (85).

Closer to the realm of intergroup relations, researchers have manipulated norm perception to change attitudes toward the outgroup (86). In one study, European American undergraduate students were asked to report their estimates of (i) the percentage of African Americans who possessed a series of stereotypical traits and (ii) the percentage of fellow students who believed African Americans to possess these traits. Afterward, if students were told that more fellow students endorsed positive stereotypes of African Americans than they had previously estimated, they subsequently reported greater endorsement of positive stereotypes; the same effect was found for negative stereotypes. In another study (86), if students received information that more other students at their own college endorsed positive stereotypes of African Americans than they had previously estimated, they subsequently reported warmer feelings (i.e. higher ratings on the feelings thermometer) toward African Americans, compared with students who received information that more students at another college had endorsed positive stereotypes. That is, it was the norm of the ingroup ("my college"), not the norm of an outgroup ("another college"), that evoked attitude assimilation. This pattern of results also reinforces our argument that ingroup- and outgroup-related perceptions are different processes that can produce different outcomes.

That these ingroup norm perception interventions have successfully shifted even persistent racial stereotypes and attitudes is promising for its application to affective polarization. The findings summarized above constitute only a small subset of the promising results from many norm perception interventions that have effectively changed diverse attitudes and behaviors in a wealth of domains (79, 87). The reason they work is that norm perceptions are dynamic. When individuals receive more accurate normative information, they update their norm perception, and subsequently shift their own beliefs to match their new perceived norm (88). Building on this process, how do we design effective norm perception interventions to curb affective polarization?

How to design and evaluate interventions that leverage ingroup norm perception to curb affective polarization

Informed by our review and analysis of the existing interventions, we suggest that the most promising approach to affective polarization should be to directly tackle partisans' perception of their ingroup's normative levels of outgroup bias. A general reason for our suggestion is that drawing attention to social norms can increase prosocial behavior (22). A more specific reason is that prior work has found that partisans dislike evidence of their ingroup displaying bias against others and that such evidence leads partisans to subsequently distance themselves from the ingroup (60, 89, 90). This occurs because group members are motivated to view their ingroup positively, as their attitudes toward their ingroups are intricately tied to their views of themselves, so they do not want to associate themselves closely with an ingroup that behaves negatively (11). As a corollary, partisans should be motivated to welcome information that their political ingroup is not as biased as they previously believed. To the extent that they incorporate the corrective information into their own belief and update their norm perception, it should produce an assimilative shift in their personal attitude toward the outgroup.

The scaffold of such interventions already exists. Experimentally testing a corrective ingroup norm perception intervention for curbing affective polarization could be as simple as flipping the manipulation of existing outgroup meta-perception interventions. Most existing interventions that correct outgroup meta-perception have been methodologically similar. Participants answer the dependent variable as a typical outgroup member, then receive the average outgroup member's actual answers, and then answer the dependent variable for themselves. Adapting the same basic structure, as illustrated in our pilot experiment (Section 1c), interventions that correct ingroup norm perception can have participants answer the dependent variable as a typical ingroup member, then receive the average ingroup member's actual answers, and then answer the dependent variable for themselves.

Using the same basic structure to correct both ingroup norm perception and outgroup meta-perception comes with an important advantage. It allows for direct comparison of the causal effects of correcting the two kinds of perception against each other, much as we have pitted their predictive effects in Sections 1a and 1b. For instance, a 2 (correcting vs. not correcting ingroup norm perception) × 2 (correcting vs. not correcting outgroup meta-perception) between-participant design would allow comparison of the main effects of correcting the two kinds of perception and detection of their potential interactions, answering a range of empirical questions: is correcting ingroup norm perception more effective than correcting outgroup meta-perception, paralleling the considerably stronger predictive effect of ingroup norm perception than of outgroup meta-perception? Or is correcting outgroup meta-perception more effective than correcting ingroup norm perception, perhaps because outgroup meta-perception is more exaggerated in the first place, giving it more room for correction (30)? Is correcting either kind of perception sufficient for curbing affective polarization such that "1 + 1 = 1 rather than 2" (26)? Or do corrections of the two kinds of perception exert independent effects ("1 + 1 = 2")?

These experimental designs can be deployed not only in artificial lab or online surveys, but also in naturalistic field settings. For an in vivo field intervention on social media, researchers could run advertisements inviting users to guess either their ingroup's feelings toward the outgroup (ingroup norm perception) or their outgroup's feelings toward the ingroup (outgroup meta-perception), and then provide correct information. By inviting users to make their best guesses about interesting or important facts, researchers could gamify the experience to maximize user interest, participation, engagement, and data quality. Users could also be incentivized (e.g. with monetary rewards) for accuracy. Previous studies have not incentivized outgroup meta-perception, but doing so in future studies could help elicit more truthful or accurate outgroup meta-perception and ingroup norm perception, providing better estimation of their effects.

The caveat is that for interventions to work in the intended direction, participants would need to see evidence that their ingroup does not actually feel as negatively toward the outgroup as they thought. If that is factually untrue, their existing negative norm perception is likely to prevail, especially if they are constantly exposed to polarized content on social media (91). But if that is

factually true, then empirical observations of less ingroup animosity toward the outgroup should effectively change norm perceptions (92), again highlighting norm perception as a promising breakpoint for intervention.

To give users further evidence of lesser polarization or to deliver additional waves of intervention, other social media advertisements could be employed to demonstrate successful intergroup interactions between political groups. The motivation for doing so is that learning about other ingroup members' successful intergroup interactions reduces one's own animosity toward the political outgroup (74, 75). Alternatively, once users' norm perception estimates are corrected, a follow-up screen could encourage users to recall their own experiences of interacting with a political outgroup member that was more positive than they expected. Such social interactions tend to be positive (93), more so than people anticipated (94, 95), giving hope for the successful recall of positive cross-party interactions in real life, which would reinforce the intervention.

Finally, turning to the dependent variable, affective polarization could be evaluated using multiple operationalizations. Feelings thermometer is an obvious possibility. Engagement with or disengagement from the political outgroup's content could serve as a proxy for contact acceptance or avoidance (96). Natural language processing techniques such as topic modeling and sentiment analysis could assess the substance of such content for mention of the political outgroup, around what themes, with what affective content and valence (61, 97, 98). These methods lend themselves well to online settings, especially social media, in which affective polarization is so rampant that simple, scalable, and effective norm perception interventions are urgently needed.

Theory-based boundary conditions that deserve prioritized research attention

A full scientific understanding of how ingroup norm perception interventions curb affective polarization requires identifying the boundary conditions. In the following, we specify the most important boundary conditions that we expect based on known theoretical properties of ingroup processes, social identity, and intergroup relations. Each boundary condition implies an extension or a limitation of our argument that deserves prioritized empirical attention. These include strength of ingroup identification, power dynamics of intergroup context, and generalizability to non-American political realities.

The effectiveness of ingroup norm perception interventions might vary as a function of the strength with which an individual identifies with their partisan ingroup. What exact patterns of moderation should be expected, however, is not straightforward. Some findings lead to the prediction that ingroup norm perception interventions should be more effective for individuals who weakly identify with their partisan ingroup; others lead to the opposite prediction.

On the one hand, ingroup norm perception interventions might be more effective among weak ingroup identifiers, paralleling other interventions against affective polarization that have been found more effective among those with weaker partisan identification. For example, intergroup contact effectively reduces affective polarization for Danish participants who weakly identify with their party but not for those who strongly identify with their party (99), presumably because strong ingroup identifiers are more motivated to view the outgroup negatively and less likely to change their attitudes toward the outgroup.

On the other hand, ingroup norm-based interventions might be more effective among strong ingroup identifiers (23) because strong ingroup identifiers are more likely to follow perceived norms of the political ingroup's issue positions (56). Moreover, the unique psychological power of political ingroup norm perception is rooted in the political ingroup being a social identity group (i.e. the political ingroup being incorporated into one's sense of self). When the political ingroup is a more relevant or important social identity, an intervention leveraging the political ingroup's norms should be more effective. Strong ingroup identifiers should value the ingroup's opinions more and be more motivated to match their attitudes to the corrective norm information and show a larger intervention effect.

Beyond the personal variable of ingroup identification strength, a contextual variable we consider important is the nature of power dynamics between one's political ingroup and outgroup. Political groups are often construed as engaging in zero-sum competitions (19, 100), which have been argued to drive negatively exaggerated outgroup meta-perception (26). But there are many naturalistic situations where intergroup cooperation is necessary, as when one party does not have enough power to push through crucial legislation and needs to shake hands across the aisle. Whether ingroup norm perception interventions have larger effects in competitive or cooperative contexts is unclear. Likewise, whether outgroup meta-perception interventions have larger effects in competitive or cooperative contexts is also unclear, because the evidence virtually always comes from competitive political contexts (26, 76, 101), and to our knowledge, there is no relevant study that has systematically compared effects in competitive vs. cooperative contexts, though some existing data should allow for such comparisons (32).

An even broader source of contextual variation is cross-national differences in political systems and climates. Much of the scholarship on affective polarization has focused, theoretically and empirically, on American politics, but the phenomenon of affective polarization is evident in many political systems (102–104), not just in the United States. We recognize that the American political context might be convenient for research on this topic both because there are only 2 major parties (so political ingroup and outgroup are easily identified) and because partisan hatred and toxicity currently run high (1). These attributes, while handy, can be a cause for scientific concern. The American political system that heightens the sense of zero-sum competition between two dominant parties might engender a unique flavor of affective polarization, sustained by powerful intergroup forces between "my allies" in a single ingroup vs. "my enemies" in a single outgroup, that is different from what is found in other political systems or realities, such as countries with multiple major parties, countries with less political gridlock and sectarianism (105), more authoritarian regimes, or places going through structural political changes.

Do the ingroup norm perception interventions that work in the unique American political context also work elsewhere? For example, do they work in countries where partisan identity is not the most relevant or salient political identity? Do they work in multiparty systems, in which the political ingroup could permeate party boundaries (106) and in which it is less clear whether all parties other than one's own are political outgroups? Future work should strategically identify political contexts that vary on these parameters and test the extent to which ingroup norm perception interventions are effective across such contexts. The goal of this empirical strategy is to use limited resources (e.g. researchers' time and money) wisely to generate evidence that is maximally

informative about the generalizability of intervention effectiveness across a vast theoretical space.

Conclusions

Throughout this Perspective piece, we have highlighted the power of ingroup norm perception in driving affective polarization. Evidence from 3 studies supports this central idea. Partisans' perception of their political ingroup's normative attitudes toward their political outgroup is a strong predictor of their own attitudes toward the political outgroup. Ingroup norm perception has more robust predictive effects on polarization-related outcomes and explains considerably more variance than does outgroup metaperception. Correcting ingroup norm perception can improve attitudes toward the political outgroup.

The unique impact of ingroup norm perception is grounded in social and psychological forces that have been well established in other group domains. Partisans generally perceive that the average ingroup member feels more negatively toward the political outgroup than they themselves feel. Conforming to this negatively exaggerated ingroup norm perception, partisans become more affectively polarized.

Existing explanations of and interventions against affective polarization have drawn on some aspects of the political ingroup but have not explicitly recognized the power of ingroup norm perception. Our critical appraisal of existing explanations suggests that they could be construed through the lens of normative processes. Our survey of existing interventions suggests that ingroup norm perception interventions have been found effective for changing a multitude of attitudes and behaviors, even sticky ones such as racial attitudes and energy-conserving behaviors. These observations reinforce our argument that correcting norm perception is a promising approach for curbing affective polarization. The stage is set for designing, testing, and evaluating such interventions in the field.

If they work, there can be further downstream benefits of incorporating ingroup norm perception into strategies for changing broader attitudes toward political others. For example, considering that affective polarization predicts ideological polarization and cross-partisan competition/cooperation (2, 3), changing ingroup norm perception holds the promise for directly improving these outcomes or indirectly improving them through curbing affective polarization. Such strategies would echo prior success in tackling sacred values underlying international conflicts and negotiations by making symbolic concessions that reduce toxic affective intensity (107).

To conclude, much work on political attitudes, including affective polarization, is built on the theoretical premise that political groups invoke social identity. That means they should display psychological properties common to social identity groups in other domains. One of the most robust properties is that fellow ingroup members' attitudes are vital to the formation of one's own attitudes. Prevailing scientific explanations of affective polarization, without properly incorporating ingroup norms, are at best an incomplete perspective and at worst a failure to recognize the impact of crucial mental processes. We urge scholars and practitioners to stop ignoring the power of ingroup norm perception in explaining and curbing affective polarization.

Materials and methods

Our original study and pilot experiment were approved by the research ethics board at the University of Toronto (RIS Human Protocol #42378). All participants provided informed consent.

Acknowledgments

The authors thank James Jiang and Nicole Velev for research assistance and Noah Laskey for feedback.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This work was supported by the John Templeton Foundation (Applied Research on Intellectual Humility to S.W.S.L.); the Social Sciences and Humanities Research Council of Canada (Insight Development Grant 430-2021-00060 and Insight Grant 435-2017-0127 to S.W.S.L.); the Ontario Ministry of Research, Innovation and Science (Early Researcher Award—Round 13 to S.W.S.L.); and the Province of Ontario (Ontario Graduate Scholarship 2023 to Z.T.Y.).

Author Contributions

Both authors contributed to conceptualization, investigation, methodology, formal analysis, and project administration. Z.T.Y. contributed to data curation, visualization, and writing—original draft. S.W.S.L. contributed to resources, supervision, validation, and writing—review and editing.

Data Availability

Data, survey materials, and analysis code associated with our original data and pilot experiment are available at https://osf.io/ahy62/. Data and code for Mernyk et al., study 1 (30, 31), were deposited by the original authors at https://osf.io/63jfb/.

References

1 Finkel EJ, *et al.* 2020. Political sectarianism in America. *Science*. 370(6516):533–536.
2 Iyengar S, Lelkes Y, Levendusky M, Malhotra N, Westwood SJ. 2019. The origins and consequences of affective polarization in the United States. *Annu Rev Polit Sci*. 22(1):129–146.
3 Druckman JN, Klar S, Krupnikov Y, Levendusky M, Ryan JB. 2021. How affective polarization shapes Americans' political beliefs: a study of response to the COVID-19 pandemic. *J Exp Polit Sci*. 8(3):223–234.
4 Abramowitz AI, Webster S. 2016. The rise of negative partisanship and the nationalization of U.S. elections in the 21st century. *Electoral Stud*. 41:12–22.
5 Whitt S, *et al.* 2021. Tribalism in America: behavioral experiments on affective polarization in the Trump era. *J Exp Polit Sci*. 8(3):247–259.
6 Chen MK, Rohla R. 2018. The effect of partisanship and political advertising on close family ties. *Science*. 360(6392):1020–1024.
7 Frimer JA, Skitka LJ, Motyl M. 2017. Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *J Exp Soc Psychol*. 72:1–12.
8 Frimer JA, Skitka LJ. 2020. Are politically diverse thanksgiving dinners shorter than politically uniform ones? *PLoS One*. 15(10):e0239988.
9 Hodson G, Meleady R. 2023. Ideologically-based contact avoidance during a pandemic: Blunt or selective distancing from 'others'? *Eur J Soc Psychol*. 53(5):823–845.

10 Nam HH, Jost JT, Van Bavel JJ. 2013. Not for all the tea in China!" Political ideology and the avoidance of dissonance-arousing situations. *PLoS One*. 8(4):e59837.
11 Tajfel H, Turner JC. 1979. The social identity theory of intergroup behaviour. In: Worchel S, Austin WG, editors. *Psychology of intergroup relations*. Monterey (CA): Brooks/Cole Publishing Company. p. 7–24.
12 Huddy L, Bankert A. 2017. Political partisanship as a social identity. In: *Oxford research encyclopedia of politics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190228637.013.250
13 Abramowitz AI, Saunders KL. 2006. Exploring the bases of partisanship in the American electorate: social identity vs. ideology. *Polit Res Quart*. 59(2):175–187.
14 Greene S. 1999. Understanding party identification: a social identity approach. *Polit Psychol*. 20(2):393–403.
15 Iyengar S, Sood G, Lelkes Y. 2012. Affect, not ideology: a social identity perspective on polarization. *Public Opin Q*. 76(3): 405–431.
16 Tajfel H, Billig MG, Bundy RP, Flament C. 1971. Social categorization and intergroup behaviour. *Eur J Soc Psychol*. 1(2):149–178.
17 Olivola CY, Sussman AB, Tsetsos K, Kang OE, Todorov A. 2012. Republicans prefer republican-looking leaders: political facial stereotypes predict candidate electoral success among right-leaning voters. *Soc Psychol Person Sci*. 3(5):605–613.
18 Rule NO, Ambady N. 2010. Democrats and republicans can be differentiated from their faces. *PLoS One*. 5(1):e8733.
19 Davidai S, Ongis M. 2019. The politics of zero-sum thinking: the relationship between political ideology and the belief that life is a zero-sum game. *Sci Adv*. 5(12):eaay3761.
20 Brewer MB. 1999. The psychology of prejudice: ingroup love and outgroup hate? *J Soc Issues*. 55(3):429–444.
21 Sherif M, Sherif CW. 1953. *Groups in harmony and tension; an integration of studies on intergroup relations*. New York (NY): Harper.
22 Krupka E, Weber RA. 2009. The focusing and informational effects of norms on pro-social behavior. *J Econ Psychol*. 30(3): 307–320.
23 Terry DJ, Hogg MA. 1996. Group norms and the attitude-behavior relationship: a role for group identification. *Pers Soc Psychol Bull*. 22(8):776–793.
24 Lees J, Cikara M. 2021. Understanding and combating misperceived polarization. *Philos Trans R Soc Lond B Biol Sci*. 376(1822): 20200143.
25 Moore-Berg SL, Hameiri B, Bruneau E. 2020. The prime psychological suspects of toxic political polarization. *Curr Opin Behav Sci*. 34:199–204.
26 Lees J, Cikara M. 2020. Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nat Hum Behav*. 4(3):279–286.
27 Banda KK, Cluverius J. 2018. Elite polarization, party extremity, and affective polarization. *Electoral Stud*. 56:90–101.
28 Thompson CG, Kim RS, Aloe AM, Becker BJ. 2017. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic Appl Soc Psych*. 39(2): 81–90.
29 Cohen J, Cohen P, West SG, Aiken LS. 2002. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd ed. New York (NY): Routledge.
30 Mernyk JS, Pink SL, Druckman JN, Willer R. 2022. Correcting inaccurate metaperceptions reduces Americans' support for partisan violence. *Proc Natl Acad Sci U S A*. 119(16):e2116851119.
31 Mernyk JS, Pink SL, Druckman JN, Willer R. Study 1. OSF. Deposited 2022 [accessed 2024 Feb 29]. https://osf.io/63jfb
32 Ruggeri K, *et al.* 2021. The general fault in our fault lines. *Nat Hum Behav*. 5(10):1369–1380.
33 Morris MW, Hong YY, Chiu CY, Liu Z. 2015. Normology: integrating insights about social norms to understand cultural dynamics. *Organ Behav Hum Decis Process*. 129:1–13.
34 Crandall CS, Eshleman A, O'Brien L. 2002. Social norms and the expression and suppression of prejudice: the struggle for internalization. *J Pers Soc Psychol*. 82(3):359–378.
35 Christ O, *et al.* 2014. Contextual effect of positive intergroup contact on outgroup prejudice. *Proc Natl Acad Sci U S A*. 111(11):3996–4000.
36 Hogg MA, Turner JC, Davidson B. 1990. Polarized norms and social frames of reference: a test of the self-categorization theory of group polarization. *Basic Appl Soc Psych*. 11(1):77–100.
37 McGarty C, Turner JC, Hogg MA, David B, Wetherell MS. 1992. Group polarization as conformity to the prototypical group member. *Br J Soc Psychol*. 31(1):1–19.
38 Dorison CA, Minson JA, Rogers T. 2019. Selective exposure partly relies on faulty affective forecasts. *Cognition*. 188:98–107.
39 Lau T, Morewedge CK, Cikara M. 2016. Overcorrection for social-categorization information moderates impact bias in affective forecasting. *Psychol Sci*. 27(10):1340–1351.
40 Norris CJ, Dumville AG, Lacy DP. 2011. Affective forecasting errors in the 2008 election: underpredicting happiness. *Polit Psychol*. 32(2):235–249.
41 Derreumaux Y, Elder J, Hughes B. The influence of group norms and affective polarization on expressed ideology; 2022 [accessed 2022 Sept 8]. https://psyarxiv.com/4m9b2/
42 Cohen GL. 2003. Party over policy: the dominating impact of group influence on political beliefs. *J Pers Soc Psychol*. 85(5): 808–822.
43 Converse P. 1964. The nature of belief systems in mass publics. In: Apter DE, editor. *Ideology and discontent*. New York (NY): Free Press. p. 206–261.
44 Malka A, Lelkes Y. 2010. More than ideology: conservative–liberal identity and receptivity to political cues. *Soc Just Res*. 23 (2–3):156–188.
45 Ehret PJ, Van Boven L, Sherman DK. 2018. Partisan barriers to bipartisanship: understanding climate policy polarization. *Soc Psychol Personal Sci*. 9(3):308–318.
46 Gerber AS, Rogers T. 2009. Descriptive social norms and motivation to vote: everybody's voting and so should you. *J Polit*. 71(1): 178–191.
47 Potoczek A, *et al.* 2023. Walk this way: ingroup norms determine voting intentions for those who lack sociopolitical control. *Pers Soc Psychol Bull*. 49(5):692–708.
48 Bougher LD. 2017. The correlates of discord: identity, issue alignment, and political hostility in polarized America. *Polit Behav*. 39(3):731–762.
49 Amira K, Wright JC, Goya-Tocchetto D. 2021. In-group love versus out-group hate: which is more important to partisans and when? *Polit Behav*. 43(2):473–494.
50 Lelkes Y, Westwood SJ. 2017. The limits of partisan prejudice. *J Polit*. 79(2):485–501.
51 Lelkes Y. 2018. Affective polarization and ideological sorting: a reciprocal, albeit weak, relationship. *Forum*. 16(1):67–79.
52 Levendusky MS. 2009. The microfoundations of mass polarization. *Polit Anal*. 17(2):162–176.
53 Merkley E. 2022. Polarization eh? Ideological divergence and partisan sorting in the Canadian mass public. *Public Opin Q*. 86(4):932–943.

54 Mason L. 2015. "I disrespectfully agree": the differential effects of partisan sorting on social and issue polarization. *Am J Pol Sci*. 59(1):128–145.
55 Mason L. 2016. A cross-cutting calm: how social sorting drives affective polarization. *Public Opin Q*. 80(S1):351–377.
56 Bakker BN, Lelkes Y, Malka A. 2020. Understanding partisan cue receptivity: tests of predictions from the bounded rationality and expressive utility perspectives. *J Polit*. 82(3):1061–1077.
57 Hmielowski JD, Hutchens MJ, Beam MA. 2020. Asymmetry of partisan media effects? Examining the reinforcing process of conservative and liberal media with political beliefs. *Polit Commun*. 37(6):852–868.
58 Lelkes Y, Sood G, Iyengar S. 2017. The hostile audience: the effect of access to broadband internet on partisan affect. *Am J Polit Sci*. 61(1):5–20.
59 Levendusky MS. 2013. Partisan media exposure and attitudes toward the opposition. *Polit Commun*. 30(4):565–581.
60 Druckman JN, Gubitz SR, Levendusky MS, Lloyd AM. 2019. How incivility on partisan media (de)polarizes the electorate. *J Polit*. 81(1):291–295.
61 Wilson AE, Parker VA, Feinberg M. 2020. Polarization in the contemporary political and media landscape. *Curr Opin Behav Sci*. 34:223–228.
62 Levendusky MS, Malhotra N. 2016. Does media coverage of partisan polarization affect political attitudes? *Polit Commun*. 33(2):283–301.
63 Allport GW. 1954. *The nature of prejudice*. Cambridge (MA): Addison-Wesley Publishing Company.
64 Brown R, Hewstone M. 2005. An integrative theory of intergroup contact. In: Zanna MP, editor. *Advances in experimental social psychology*. Vol. 37. New York (NY): Elsevier Academic Press. p. 255–343.
65 Bakshy E, Messing S, Adamic LA. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science*. 348(6239):1130–1132.
66 Klar S, Krupnikov Y, Ryan JB. 2018. Affective polarization or partisan disdain? *Public Opin Q*. 82(2):379–390.
67 Boss H, Buliga E, MacInnis CC. 2023. "Everybody's doing it": exploring the consequences of intergroup contact norms. *Group Process Intergr Relat*. 26(6):1205–1222.
68 Binder J, *et al*. 2009. Does contact reduce prejudice or does prejudice reduce contact? A longitudinal test of the contact hypothesis among majority and minority groups in three European countries. *J Pers Soc Psychol*. 96(4):843–856.
69 Gaertner SL, Dovidio JF, Anastasio PA, Bachman BA, Rust MC. 1993. The common ingroup identity model: recategorization and the reduction of intergroup bias. *Eur Rev Soc Psychol*. 4(1):1–26.
70 Levendusky MS. 2018. Americans, not partisans: can priming American national identity reduce affective polarization? *J Polit*. 80(1):59–70.
71 Amsalem E, Merkley E, Loewen PJ. 2022. Does talking to the other side reduce inter-party hostility? Evidence from three studies. *Polit Commun*. 39(1):61–78.
72 Gaertner SL, *et al*. 1999. Reducing intergroup bias: elements of intergroup cooperation. *J Pers Soc Psychol*. 76(3):388–402.
73 Warner BR, Villamil A. 2017. A test of imagined contact as a means to improve cross-partisan feelings and reduce attribution of malevolence and acceptance of political violence. *Commun Monogr*. 84(4):447–465.
74 Wojcieszak M, Warner BR. 2020. Can interparty contact reduce affective polarization? A systematic test of different forms of intergroup contact. *Polit Commun*. 37(6):789–811.
75 Huddy L, Yair O. 2021. Reducing affective polarization: warm group relations or policy compromise? *Polit Psychol*. 42(2):291–309.
76 Landry AP, Schooler JW, Willer R, Seli P. 2023. Reducing explicit blatant dehumanization by correcting exaggerated meta-perceptions. *Soc Psychol Personal Sci*. 14(4):407–418.
77 Druckman JN, Klar S, Krupnikov Y, Levendusky MS, Ryan JB. 2022. (Mis)estimating affective polarization. *J Polit*. 84(2):1106–1117.
78 Dimant E, Gesche T. 2023. Nudging enforcers: how norm perceptions and motives for lying shape sanctions. *PNAS Nexus*. 2(7):pgad224.
79 Tankard ME, Paluck EL. 2016. Norm perception as a vehicle for social change. *Soc Issues Policy Rev*. 10(1):181–211.
80 Prentice DA, Miller DT. 1993. Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm. *J Pers Soc Psychol*. 64(2):243–256.
81 LaBrie JW, *et al*. 2013. RCT of web-based personalized normative feedback for college drinking prevention: are typical student norms good enough? *J Consult Clin Psychol*. 81(6):1074–1086.
82 Su J, *et al*. 2018. Evaluating the effect of a campus-wide social norms marketing intervention on alcohol-use perceptions, consumption, and blackouts. *J Am Coll Health*. 66(3):219–224.
83 Allcott H, Mullainathan S. 2010. Behavior and energy policy. *Science*. 327(5970):1204–1205.
84 Nolan JM, Schultz PW, Cialdini RB, Goldstein NJ, Griskevicius V. 2008. Normative social influence is underdetected. *Pers Soc Psychol Bull*. 34(7):913–923.
85 Cookson D, Jolley D, Dempsey RC, Povey R. 2021. A social norms approach intervention to address misperceptions of anti-vaccine conspiracy beliefs amongst UK parents. *PLoS One*. 16(11):e0258985.
86 Stangor C, Sechrist GB, Jost JT. 2001. Changing racial beliefs by providing consensus information. *Pers Soc Psychol Bull*. 27(4):486–496.
87 Cialdini RB, Goldstein NJ. 2004. Social influence: compliance and conformity. *Annu Rev Psychol*. 55(1):591–621.
88 Paluck EL, Shepherd H. 2012. The salience of social referents: a field experiment on collective norms and harassment behavior in a school social network. *J Pers Soc Psychol*. 103(6):899–915.
89 Frimer JA, Skitka LJ. 2018. The Montagu principle: incivility decreases politicians' public approval, even with their political base. *J Pers Soc Psychol*. 115(5):845–866.
90 Shafranek RM. 2020. Political consequences of partisan prejudice. *Polit Psychol*. 41(1):35–51.
91 Brady WJ, *et al*. 2023. Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility. *Nat Hum Behav*. 7(6):917–927. https://doi.org/10.1038/s41562-023-01582-0
92 Bicchieri C, Mercier H. 2014. Norms and beliefs: how change occurs. In: Xenitidou M, Edmonds B, editors. *The complexity of social norms, computational social sciences*. New York (NY): Springer International Publishing. p. 37–54.
93 Page-Gould E. 2012. To whom can I turn? Maintenance of positive intergroup relations in the face of intergroup conflict. *Soc Psychol Personal Sci*. 3(4):462–470.
94 Schroeder J, Lyons D, Epley N. 2022. Hello, stranger? Pleasant conversations are preceded by concerns about starting one. *J Exp Psychol: Gen*. 151(5):1141.
95 Wald KA, Kardas M, Epley N. 2024. Misplaced divides? Discussing political disagreement with strangers can be unexpectedly positive. *Psychol Sci*. 35(5):471–488.
96 Le Forestier JM. 2023. *Prejudice reduction through intergroup contact on social media*. Toronto (ON): University of Toronto.

97 Wang SN, Inbar Y. 2021. Moral-language use by U.S. political elites. *Psychol Sci*. 32(1):14–26.
98 Frimer JA, *et al*. 2023. Incivility is rising among American politicians on Twitter. *Soc Psychol Personal Sci*. 14(2): 259–269.
99 Thomsen JPF, Thomsen AH. 2023. Intergroup contact reduces affective polarization but not among strong party identifiers. *Scand Polit Stud*. 46(4):241–263.
100 Roberts R, Davidai S. 2022. The psychology of asymmetric zero-sum beliefs. *J Pers Soc Psychol*. 123(3):559–575.
101 Moore-Berg SL, Ankori-Karlinsky L-O, Hameiri B, Bruneau E. 2020. Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proc Natl Acad Sci U S A*. 117(26):14864–14872.
102 Bantel I. 2023. Camps, not just parties. The dynamic foundations of affective polarization in multi-party systems. *Electoral Stud*. 83:102614.
103 Cochrane C. 2010. Left/right ideology and Canadian politics. *Can J Polit Sci*. 43(3):583–605.
104 Lachat R. 2008. The impact of party polarization on ideological voting. *Electoral Stud*. 27(4):687–698.
105 The Economist Intelligence Unit. 2024. *Democracy index 2023: age of conflict*. London: The Economist Group.
106 You ZT, Page-Gould E, Thai S, Le Forestier JM. The psychological boundaries of political groups; 2024 [accessed 2024 Mar 29]. https://osf.io/5qvea
107 Atran S, Axelrod R. 2008. Reframing sacred values. *Negotiation J*. 24(3):221–246.