

# Overestimating the Social Costs of Political Belief Change

[Trevor Spelman](#)

Northwestern University and IPR

[Abdo Elnakouri](#)

Northwestern University

[Nour Kteily](#)

Northwestern University

[Eli Finkel](#)

Northwestern University and IPR

Version: September 16, 2024

**DRAFT**

*Please do not quote or distribute without permission.*

## Abstract

How do U.S. partisans expect members of their political ingroup to react when they shift away from the typical view of their party (e.g., a Democrat adopting a more conservative stance on private gun ownership)? And how well do these expectations align with actual reactions? Five studies ( $N = 2,655$ ) employing diverse research methods—including surveys, behavioral outcomes, live participant interactions, and qualitative measures—revealed that partisans' expectations are systematically and substantially miscalibrated: they overestimate how much they would be socially sanctioned for dissenting belief change with an average weighted effect size ( $d$ ) of .84. The researchers find that this overestimation is partially driven by self-protective egocentric bias and show that it decreases via an intervention that promotes perspective taking. Additional results revealed that these miscalibrated expectations predict an increased likelihood of self-censorship. By examining the subtle social forces that exert pressures toward ingroup conformity, this work offers insights into how we can foster more open political dialogue within political groups.

### Abstract

How do U.S. partisans expect members of their political ingroup to react when they shift away from the typical view of their party (e.g., a Democrat adopting a more conservative stance on private gun ownership)? And how well do these expectations align with actual reactions? Five studies ( $N = 2,655$ ) employing diverse research methods—including surveys, behavioral outcomes, live participant interactions, and qualitative measures—revealed that partisans' expectations are systematically and substantially miscalibrated: they overestimate how much they would be socially sanctioned for dissenting belief change with an average weighted effect size ( $d$ ) of .84. We find that this overestimation is partially driven by self-protective egocentric bias and show that it decreases via an intervention that promotes perspective taking. Additional results revealed that these miscalibrated expectations predict an increased likelihood of self-censorship. By examining the subtle social forces that exert pressures toward ingroup conformity, this work offers insights into how we can foster more open political dialogue within political groups.

*Keywords:* Social misperceptions; political misperceptions; dissent; self-censorship; social conformity

### Statement of Limitations

Despite the robustness and consistency of our findings across studies, several limitations should be noted. First, our research was conducted within the specific political and cultural context of the United States during a period of heightened political polarization, which may limit the generalizability of our results to other cultural or political settings. Second, our studies primarily involved exchanges between participants who were strangers or had limited prior acquaintance, which allowed us to isolate certain psychological mechanisms but may not fully capture the complexities of interactions within close relationships. Third, our research focused on dyadic exchanges, which may not reflect the dynamics of larger group interactions, particularly in online environments. Finally, while our findings demonstrate that third-party perspective taking can reduce egocentric bias and lead to more accurate social judgments, this intervention did not fully eliminate the perception gap. Future research should explore ways to strengthen self-distancing interventions and examine other factors that may contribute to this perception gap.

**Abstract Word Count: 160**

**Introduction and Discussion Sections Word Count: 4,907**

## Introduction

In an era marked by deep political divides, it might seem unlikely that American partisans would update their views on divisive issues. Despite the well-documented tendency to perceive reality according to one's political allegiances (Van Bavel & Pereira, 2018), individuals do sometimes update their beliefs—even when this means drifting away from the prevailing orthodoxy of their political party (e.g., Kossowska et al., 2023; Tappin et al., 2023; Xu & Petty, 2022). Such shifts entail risk: Individuals who dissent from this orthodoxy may face harsh reactions from other group members (e.g., Jetten & Hornsey, 2014; Kruglanski & Webster, 1991; Schachter, 1951). For example, a committed Republican who adopts a less conservative stance on abortion confronts a tension. They may disclose their evolving views at the risk of social sanction or conceal this shift to retain the status of a loyal group member.

Insofar as disclosing dissenting belief change<sup>1</sup> entails risks of social rejection, concealing this change may be a wise choice. But what if assessments of this risk are systematically miscalibrated? When the risk of rejection is salient, dissenters—sensitive to rejection and limited by their own egocentric perspective—may overestimate how negatively others in their group will respond. These exaggerated expectations can lead individuals to self-censor their views, with consequences for political discourse and the representation of ideas in the public sphere. Given that partisans tend to socialize with (Iyengar et al., 2012; Mason & Wronski, 2018), live near (Brown & Enos, 2021; Motyl et al., 2014), and discuss politics (Mutz, 2006) with ingroup

---

<sup>1</sup> We use the term “dissenting belief change” to characterize the phenomenon in which an individual's beliefs regarding a divisive political issue become less aligned with the typical view of their political party to any extent. We refer to individuals who experience such changes as “dissenters”.

members significantly more often than outgroup members, the study of social and psychological factors that influence intragroup dissent is increasingly important.

We propose that individuals are more likely to self-censor dissenting belief change insofar as they expect to be rejected by other group members for their dissent. However, we hypothesize that these expectations are systematically miscalibrated: Dissenters overestimate how much other group members will reject them for their dissent. We also hypothesize that this perception gap is partially driven by egocentric bias, which is heightened when the risk of rejection is made salient, and may be reduced through interventions that enhance perspective taking. We test these ideas across five studies. In the following sections, we ground these hypotheses in social psychological theory, review relevant studies, and highlight the novel contributions of this research.

### **The Need to Belong and Fear of Social Rejection**

The need to belong is a fundamental feature of human psychology (Baumeister & Leary, 1995). People largely define themselves by the groups they affiliate with, and group membership provides many benefits—including a sense of self-worth and access to various forms of material and social capital (Correll & Park, 2005; Gaertner et al., 2015; Hogg, 2016; Kruglanski et al., 2006; Swann et al., 2012; Van Bavel & Packer, 2021). Given the central role of group membership in human social life, people are highly averse to the threat of social rejection.

This aversion may be rooted in our evolutionary past. Throughout human history, the consequences of social rejection posed serious threats to survival (see Williams, 2007, for review). In early hunter-gatherer societies, the excommunicated were likely to die if they could not repair damaged social bonds or form new ones with another group (Baumeister & Leary, 1995; Henrich, 2015). Many scholars have argued that, in response to such threats, humans have

evolved a “better safe than sorry” approach to assessing the risk of social rejection (Baumeister & Tice, 1990; Haselton & Buss, 2000; MacDonald & Leary, 2005). Thus, we hypothesize that this risk-averse orientation will manifest in a tendency to overestimate the social costs of dissent.

These expectations may have implications for social conformity and self-expression. One recent study documents an increase in the need to belong over the past two decades as people have become increasingly reticent to openly discuss their views (Chopik et al., 2024). In political contexts, where groups use beliefs to distinguish who is and who is not part of the group (Bakker et al., 2020; Connors, 2020; Dovidio et al., 2017; Gaertner et al., 2015; Haidt, 2012; Van Bavel & Pereira, 2018), individuals may anticipate that other group members will view dissenting belief change as an act of punishable deviance that could jeopardize their standing in the group.

Recent research highlights the mechanisms through which political groups reinforce conformity and punish deviance. Studies have shown that American partisans prefer ingroup members who strongly endorse the political views of their party over those with more moderate and diverse political views (Goldenberg et al., 2023; Zimmerman et al., 2022). Additional research has shown that partisans reward group members who selectively attend to information that supports their own party’s views (Moore et al., 2023) and penalize group members who are receptive to opposing views (Hussein & Wheeler, 2024; c.f., Heltzel & Laurin, 2021)—especially when such receptiveness signals dissent (Heltzel & Laurin, 2021, Study 3). Taken together, these findings suggest that group members who dissent may be prudent to fear harsh social reactions to dissenting belief change. But what if these expectations are exaggerated?

We suggest that the risk of rejection for dissenting belief change looms larger in the mind than it does in reality. We derive this hypothesis in part from two research streams. First, people tend to overestimate how much others attend to their actions and behaviors (the spotlight effect:

Gilovich et al., 2000, 2002) and tend to see themselves as the target of others' thoughts and attention (Leary & Downs, 1995; Zuckerman et al., 1983). Second, people often underestimate how tolerant and receptive others are toward disagreement. For instance, people overestimate the negativity of political disagreement (Wald et al., 2024) and others' negative emotional reactions to dissenting views (Dorison & Minson, 2022). Aligned with the idea that group members adopt a risk-averse approach toward social rejection, these findings further suggest that people overestimate others' negative reactions to dissenting belief change, leading to a heightened fear of rejection that does not align with reality.

### **Psychological Immersion and Egocentric Bias**

We suggest that this general tendency (to overestimate how much ingroup members will reject dissenting belief change) is driven in part by egocentric bias. *Egocentric bias*—the tendency to rely heavily on one's own perspective and failing to fully consider others' points of view—limits perspective taking and contributes to interpersonal conflict (Epley et al., 2004; Gilbert, 2002; Gilovich et al., 1999; Nickerson, 1999; Tversky & Kahneman, 1974). This bias is also relevant to risk-aversion: People tend to be less willing to take social risks (e.g., asking someone out on a date; Beisswanger et al., 2003) when considering such decisions for themselves than for others (Batteux et al., 2019). Similarly, research on the risk-as-feelings hypothesis demonstrates that people focus less on the emotional consequences associated with a given risk when considering decisions for others compared to the self (Loewenstein et al., 2001). Given such evidence, we suggest that adopting a third-party perspective when estimating ingroup member reactions to dissent will reduce exaggerated expectations of rejection.

Building on the insight that people are less sensitive to risk when considering decisions for others than for themselves, we propose that adopting a third-party perspective will attenuate

egocentric bias by fostering psychological distance (e.g., Grossmann & Kross, 2014).

Psychological distancing is a mental process of creating a sense of separation between oneself and a particular event, situation, or emotion. Research has shown that adopting a psychologically distanced perspective reduces egocentric bias and contributes to greater coping with uncertainty and stress (Ayduk & Kross, 2010; Fujita et al., 2006; Golubickis et al., 2016; Grossmann et al., 2021; Grossmann & Kross, 2014; Kross & Ayduk, 2017, 2024; Trope & Liberman, 2003). One effective technique to promote psychological distance is to adopt a third-party perspective, which involves viewing a situation as if one is an outside observer rather than a participant (Libby & Eibach, 2011; Piaget, 1926). This technique is effective during stressful and conflictual social situations (Finkel et al., 2013; Kross et al., 2014, Study 2) and mitigates the self-other asymmetry in wise reasoning (Grossmann & Kross, 2014).

Given humans' better-safe-than-sorry orientation toward social rejection (Baumeister & Tice, 1990; MacDonald & Leary, 2005), we hypothesize that dissenters—sensitive to the potentially debilitating costs of social rejection they might experience—will overestimate social rejection as a consequence of egocentric bias. Moreover, we propose that dissenters will make more accurate social judgments about ingroup member reactions when egocentric bias is reduced. We hypothesize that adopting a third-party point of view will promote psychological distancing and reduce egocentric bias, thereby narrowing the gap between predicted and actual ingroup reactions to dissenting belief change. In other words, we hypothesize that people will make more accurate predictions about ingroup social reactions when considering the situation from a psychologically distant third-party perspective compared to a psychologically immersed first-person perspective.

### **Intragroup Misperceptions**



Lastly, our work draws inspiration from previous work on group-level misperceptions in the realm of politics (Fernbach & Van Boven, 2022; Ruggeri et al., 2021) and makes two novel contributions to this literature. First, our work focuses on the causes and consequences of misperceptions within political groups. Whereas previous studies have documented the powerful effects of social misperceptions *between* political groups, the study of social misperceptions *within* political groups has received much less attention. Preliminary studies that have examined intragroup misperceptions find that they exacerbate affective polarization and ideological extremity (Brady et al., 2023; Lees & Cikara, 2020), thus highlighting the importance of future investigation in this area. We contribute to this important area of work by studying the causes and consequences of one type of intragroup misperception.

Second, our work introduces a novel approach to correcting misperceptions. Previous studies on intergroup misperceptions have shown that providing accurate factual information about the outgroup can effectively reduce misperceptions (Ruggeri et al., 2021; Lees & Cikara, 2020; Mernyk et al., 2022; c.f., Dias et al., 2024; Nyhan & Reifler, 2010). We extend this line of research by exploring how addressing the psychological source of misperceptions (in this case, egocentric bias) helps to produce more accurate social judgments. Thus, whereas previous studies have adopted an informational approach to correcting group-level misperceptions, our studies explore whether a psychological approach—focusing on the underlying psychology that creates misperceptions—may be similarly effective.

### **Hypotheses and Research Overview**

Based on the preceding analysis, we advance three central hypotheses in the present research:

*H1: Partisans will be increasingly likely to self-censor dissenting belief change insofar as they fear social rejection from other group members for dissent.*

*H2: Partisans will systematically overestimate how much other ingroup members will reject them for dissenting belief change.*

*H3: Adopting a third-party perspective will increase psychological distancing and produce more accurate predictions about social rejection.*

We tested these hypotheses across 5 studies ( $N = 2,655$ ; four pre-registered; three additional supplemental studies) using behavioral outcomes, paired dyads, and live interactions. First, we examine how expectations about ingroup member reactions influence group members' decisions to disclose versus conceal dissenting political belief change (Pilot Study 1, Study 2). Second, we test whether these expectations—across varied degrees of belief change and relational familiarity between individuals—align with how ingroup members actually react (Studies 1-5). Finally, we examine how egocentric bias contributes to inaccurate perceptions about ingroup reactions and explore psychological distancing as an intervention to attenuate such bias and correct misperceptions (Study 5).

### **Transparency and Openness**

The design, hypotheses, and analysis plan for studies 2, 3, 4, and 5 were pre-registered, which may be accessed using the link and password in the footnote below<sup>2</sup>. Any deviations from the pre-registered document are reported in Table S1 in the Supplementary Online Materials (SOM; according to guidelines by Willroth & Atherton, 2024). To facilitate accessibility, we provide access to all experimental measures, procedures, statistical analyses, data, and pre-

---

<sup>2</sup> Link: [https://researchbox.org/3336&PEER\\_REVIEW\\_passcode=NBUYVQ](https://researchbox.org/3336&PEER_REVIEW_passcode=NBUYVQ)  
Password: NBUYVQ

registrations at the same link. Our studies used varied samples that included participants from the online participant recruitment platforms Amazon Mechanical Turk (via CloudResearch) and Prolific. For all studies, we report how we determined our target sample size, all data exclusions, all manipulations, and all measures in the methods sections. This research was approved by the [blinded for peer-review] University Ethics Board. ChatGPT was used to edit code for data analysis, revise select sentences in the manuscript, and conduct basic literature reviews in the preparation of this manuscript (OpenAI, 2024).

### **Pilot Study 1**

Although the relationship between fear of social rejection and self-censorship has been well established in previous work (Glynn et al., 1997; Matthes et al., 2017; Moy et al., 2001; Noelle-Neumann, 1993; Scheufele et al., 2001), this relationship has not been examined specifically in the context of dissenting belief change. To examine whether this relationship is present in this context, we conducted a pilot study wherein we surveyed U.S. partisans who reported dissenting belief change in the past 12 months ( $N = 131$ ) and asked them (a) how they expect ingroup members to react when learning about their dissenting belief change<sup>3</sup>, and (b) whether they have become more or less likely to self-censor their views on this topic in the time since their belief change occurred. This design represents an ecologically valid context wherein partisans who have experienced dissenting belief change reported their real-world behaviors and expectations when interacting with other group members. We found that participant expectations

---

<sup>3</sup> Expectations about ingroup member reactions were measured on a nine-item composite measure of social sanctions, which was also used in Study 2 and is described in detail in the Study 2 Methods.

for social sanction were significantly correlated with a greater likelihood of self-censorship, Spearman's  $r = -.22$ ; two-tailed  $p = .014^4$ .

In line with previous research, participants who expected harsher social reactions from members of their ingroup for dissenting belief change were more likely to self-censor when discussing this political topic with other group members. We believe this study represents a conservative test of the relationship between anticipated rejection and self-censorship because people are often unaware of social influences on their behavior (Asch, 1956; Cialdini & Goldstein, 2004); therefore, participants in this survey likely underreported the extent to which they self-censor and fear social sanctions. Overall, these findings—which, as discussed later, replicate in Study 2—demonstrate that group members are more likely to self-censor dissenting belief change insofar as they expect harsh social reactions from other group members for doing so, a tendency that elevates the stakes of potentially distorted rejection estimates.

## Study 1

Given the relationship between anticipated social sanction and self-censorship of dissenting belief change, an important question is whether estimates about ingroup member reactions are well calibrated. In Study 1, we investigated this question by comparing predicted versus actual social rejection for dissenting belief change.

### Methods

#### *Participants*

---

<sup>4</sup> See Pilot Study 1 in the SOM for a full description of the procedure, measures, analyses, and results from this study.

We recruited 519 participants from Amazon's Mechanical Turk using custom recruitment filters on CloudResearch to achieve a sample with equal numbers of Democrat and Republican participants. We aimed for a final sample of 500 participants to detect a small effect and to test for possible interactions by participant partisanship. We conducted an a priori power analysis using G\*Power 3.1 (Faul et al., 2009), which suggested a sample size of at least 420 participants to detect a small-to-moderate effect ( $f = 0.2$ ) size with 90% power at  $\alpha = .05$ . Given the novelty of this research area and our aim to generalize findings across multiple political topics, we decided to over-recruit to enhance the precision of, and confidence in, our findings. We excluded thirteen participants who did not complete the study, one participant who failed an attention check, and five participants who reported that the study procedure was confusing. Our final sample consisted of 500 participants (248 Democrats, 252 Republicans; 270 females, 225 males, 5 self-identify;  $M_{age} = 44.95$ ;  $SD_{age} = 12.68$ ). This study was not pre-registered.

### ***Materials and Procedure***

Study 1 used a two-cell between-participants design wherein participants were randomly assigned either to a *predict* (referred to as “predictors”) or a *react* (referred to as “reactors”) condition. First, participants reported their political orientation and attitudes on three key political issues in the United States: immigration policy, abortion access, and private gun ownership. We chose these topics based on evidence that opinions on these topics are strongly polarized between political parties and largely homogenous within political parties (Hawkins et al., 2018).

Participants indicated their stance on each of these issues by selecting from binary response options representing a liberal view and a conservative view on each issue. For example, for the abortion access question, participants were asked, “Which of the following statements

best reflects your views on the issue of abortion access in the United States?” Participants then selected either, “Abortion access in this country should be **protected**” (liberal view), or “Abortion access in this country should be **restricted**” (conservative view).

Among the topics that participants reported a view that aligned with the typical view of their political party<sup>5</sup>, we randomly assigned participants to consider one of these topics as the focal topic for the next part of the study. For example, if a Democrat participant reported a conservative view on gun control, that topic would not have been eligible for random selection as the focal topic. This process ensured that all participants were being asked about a topic regarding which they held a view that aligned with the typical view of their party<sup>6</sup>.

Next, participants responded to dependent measures. Predictors were asked to predict how a political ingroup member would react if they adopted the opposing party’s view on the topic (i.e., dissenting belief change). Reactors reported their reactions to an ingroup member who adopted the opposing party’s view on the topic. For example, a Republican participant in the predict condition who reported a conservative view on gun ownership would be asked to predict how another Republican would react if they were to change their mind on the issue of gun ownership from “Access to private gun ownership in this country should be **protected**” to “Access to private gun ownership in this country should be **restricted**.” If this participant were in the react condition, they would be asked to report their reactions toward another Republican whom they were told adopted the opposing party’s view.

---

<sup>5</sup> Across all topics, participants in this sample reported that they held the typical view of their party 82.4% of the time. We do not see significant differences in the results reported when controlling for whether participants hold counter-normative views. These analyses are reported in the SOM.

<sup>6</sup> We used these three political topics and this process of random assignment to political topics in Studies 3, 4, and 5. Study 2 considered only the issue of abortion access.

We used a five-item scale measure of social rejection (adapted from Cavazza et al., 2014), which served as our key dependent measure in this study ( $\alpha = .94$ ). We used this measure because social rejection is an experience that people have a strong desire to avoid (Williams, 2007) and is especially relevant in the realm of intragroup dynamics and attitude deviance (Jetten & Hornsey, 2014). The items in the predict condition included, “If you were to change your mind on the issue of [topic] to believe [the opposite] to what extent do you think that another [Democrat / Republican] would... (1) exclude you? (2) ignore your input? (3) reject you? (4) disrespect you? (5) criticize you?”. All items were answered on a Likert-type scale (1 = *Not at all*; 7 = *Very much*). The items in the react condition were similar, but the wording was modified such that reactors reported how they would react to another member of their political ingroup who changed their mind.

Finally, participants responded to attention checks and demographic questions including age, gender, ethnicity, and U.S. zip code<sup>7</sup>.

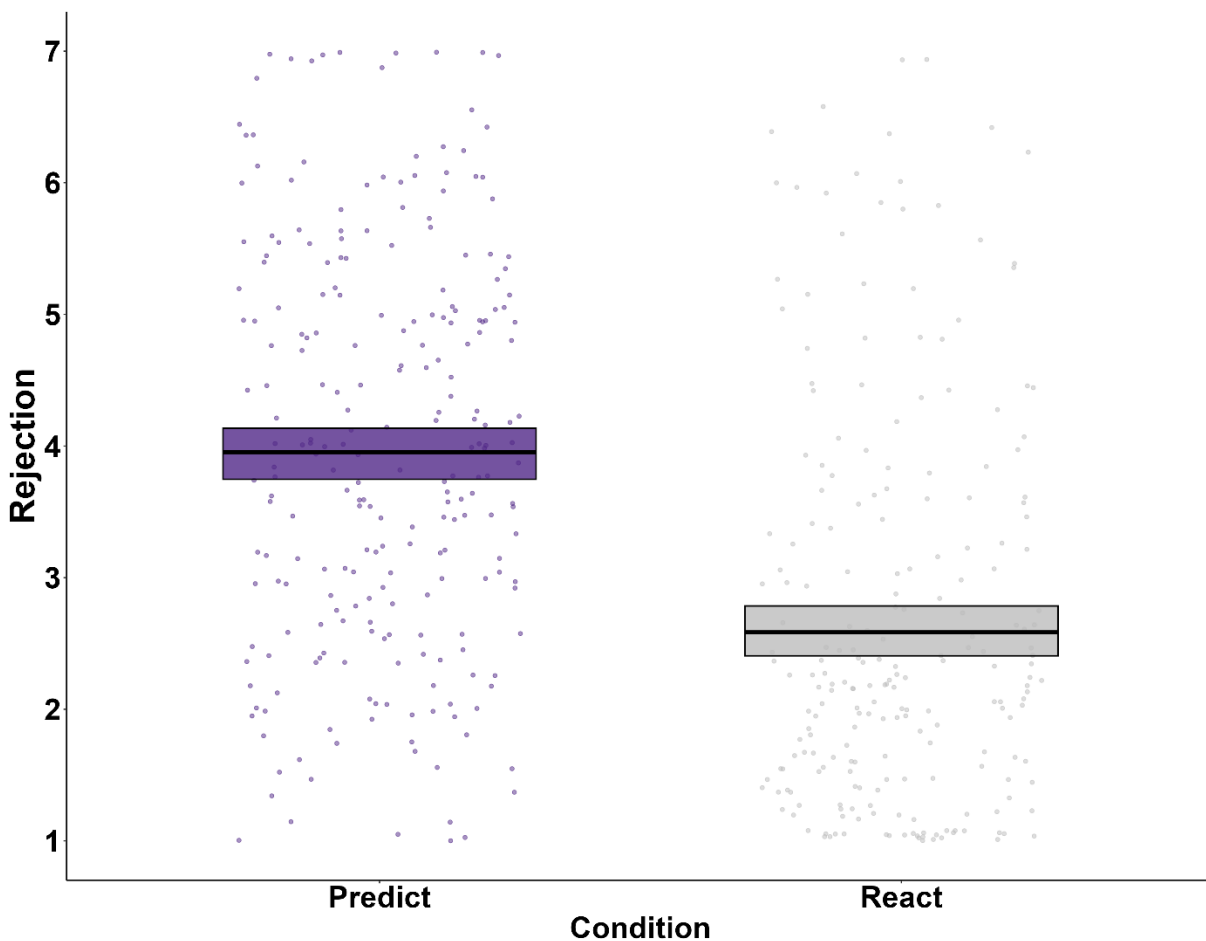
## **Results**

Were predictors’ expectations about ingroup member reactions to dissent accurate? We conducted an independent samples t-test to compare predictors’ expectations for social rejection to reactors’ actual reports. In support of our hypothesis, this test revealed that predictors anticipated significantly more social rejection for dissenting belief change ( $M = 3.95$ ,  $SD = 1.58$ ) compared to the reactions that reactors reported ( $M = 2.59$ ,  $SD = 1.59$ ),  $t(498) = 9.65$ ,  $p < .001$ ,  $d = .86$  (Figure 1). Next, we conducted a series of independent samples t-tests comparing social rejection composite scores between predict vs react role conditions for each randomly assigned

---

<sup>7</sup> Following our primary dependent measures, participants responded to additional survey items that are not included in the present research.

political topic. These tests revealed that the effect of condition was significant for each topic (all  $p$ 's < .001;  $d_{abortion} = .90$ ;  $d_{immigration} = .77$ ;  $d_{guncontrol} = .89$ ), suggesting that this pattern of overestimation extends across various political topics. Next, we conducted a two-way ANOVA to explore whether there were any significant differences between Democrat and Republican participants in our sample, which revealed that there were no significant differences between Democrats and Republicans in the effect of role on the composite measure of rejection ( $p = .502$ ), implying that the discrepancy between predicted and actual reactions was similarly present for Democrats and Republicans. The full results of the analyses for each topic and by participant party identity are reported in the SOM.





**Figure 1.** *Predictors overestimate rejection relative to reactors' reports.* Note. The data in this plot are slightly jittered to avoid overlap. The height of the boxes represents the 95% confidence interval.

### ***Discussion***

Study 1 documents a perception gap, showing that predictors significantly overestimated how much they would be rejected by a political ingroup member for dissenting belief change. This pattern was consistent across each of the three political topics examined in this study and was similarly present for both Democrat and Republican participants in our sample.

One potential limitation of this study is that participants were asked to consider an atypical type of belief change: shifting from one side to the other side on a contentious political issue (e.g., from pro-life to pro-choice). To ensure the generalizability of these findings, we investigated several types of dissenting belief change across Studies 2, 3, and 4<sup>8</sup>. A second possible limitation of Study 1 is that participants were only asked Likert-scale survey items, raising the possibility that reactors in our sample responded in a socially desirable way (i.e., wanting to appear open-minded). Therefore, in Studies 2 and 3, we included financially incentivized behavioral measures of punishment to examine whether reactors' survey-item reports are consistent with their behavior.

We now turn to exploring possible explanations for this phenomenon. Previous research shows that when people are asked to imagine the typical partisan, their representations tend to be more attitudinally extreme than partisans are in reality (e.g., Brady et al., 2023; Druckman & Levendusky, 2019; Levendusky & Malhotra, 2016; Yudkin et al., 2019). Therefore, it is

---

<sup>8</sup> See also Pilot Study 2 in the SOM, which replicated Study 1 for a minor change in beliefs.

plausible that predictors in Study 1 overestimated social rejection because they were imagining an extreme ingroup member (i.e., someone who strongly endorses the party's positions on issues), which may have heightened expectations for harsh social rejection. If predictor misperceptions about social rejection are driven by exaggerated representations of reactors' extremity, then this perception gap should disappear when predictors are provided with accurate information about reactor extremity. We conducted Study 2 as a test of this explanation.

## **Study 2**

Study 2 had three primary objectives. First, it tested whether providing predictors with accurate information about the attitudinal extremity of their react condition counterpart would lead to more accurate estimates of social sanctions for dissenting belief change. Second, it investigated whether reactors' self-reports of social sanctions are consistent with their behaviors toward dissenting ingroup members. Third, Study 2 included a measure of self-censorship among predictors to further investigate the relationship between anticipated social sanctions and self-censorship.

## **Methods**

### ***Design***

Study 2 employed a 2 (role: predict vs react)  $\times$  2 (extremity: extreme vs moderate) between-participants design and was conducted in two waves, which occurred roughly twenty-four hours apart.

### ***Participants***

Participants were recruited from Prolific. Wave 1 served as a pre-screen to identify participants to recruit for Wave 2 based on their responses to attitude extremity measures in Wave 1. Per our pre-registration, we aimed to recruit a final sample of 400 participants to complete both waves of the study. This sample size was determined by power analysis using G\*power to estimate the minimum sample size required to observe an interaction between conditions for a two-way ANOVA with an effect size of *Cohen's d* = .03 at 95% power.

During Wave 1, participants were asked to report the following: (1) Their partisan identity, (2) which of the following statements best represents their views on abortion access: “Abortion access in this country should be [protected/restricted]”, and (3) how strongly they agree with the statement that they selected on a 0-100 scale (0 = *Completely Disagree*; 100 = *Completely Agree*). We used custom recruitment filters on Prolific to recruit a sample of 2,069 Democrat and Republican participants during Wave 1 with the goal of identifying at least 240 (120 Republicans; 120 Democrats) “extreme” participants who indicated that they agreed with the statement they selected on abortion between 95-100 on the agreement scale and at least 240 (120 Republicans; 120 Democrats) “moderate” participants who indicated they agree with the statement between 50-85 to re-recruit for Wave 2<sup>9</sup>.

Participants from Wave 1 who reported that they held the typical view of their party on the topic of abortion and agreed with that view between either 50-85 or 95-100 on the attitude agreement measure were eligible to participate in Wave 2. Among these, we randomly selected an equal number of Democrats and Republicans who met our recruitment criteria in each

---

<sup>9</sup> At the end of Wave 1, participants were asked additional questions about their attitudes and the attitudes of other group members on the topic of abortion that are not pertinent to the present study.

category ( $n = 515$ ) and invited them to participate in Wave 2 twenty-four hours later<sup>10</sup>. Of these, 435<sup>11</sup> accepted our invitation to participate in Wave 2 and 401 participants met all criteria to be included in our final sample (221 Democrats, 190 Republicans; 213 females, 185 males, 3 self-identify;  $M_{age} = 43.05$ ;  $SD_{age} = 13.92$ ).

Our sample consisted of near-equal numbers of participants who agreed with the typical view of their party either between 50-85 (moderates;  $n = 200$ ) or between 95-100 (extreme;  $n = 201$ ) on the 0-100 agreement scale. We specified these attitude-based grouping criteria based on participant responses to previous studies and feasibility. Regarding the latter, we observed a high concentration of participants who reported that they strongly agreed with the typical view of their party in previous studies, whereas participants reporting more moderate views were distributed across a wider range of agreement. After conducting an initial pilot study for Wave 1, we realized that it would not be feasible to recruit a sufficient sample size of “moderate” participants if we narrowly defined the range by five scale points (e.g., 75-80), as in the “extreme” category. Therefore, we created a relatively narrower grouping category for “extreme” agreement and a broader category for “moderate” agreement. We report all data on the sample statistics for Wave 1 and Wave 2 in the SOM.

### ***Procedure***

At the beginning of Wave 2, we randomly assigned participants to either the predict or the react role condition. We told predictors that they would be paired with another participant

---

<sup>10</sup> There were no significant differences between the randomly selected subset of eligible participants from Wave 1 and those who were not randomly selected for participation in Wave 2 in terms of their agreement with the statement or their political identity centrality. We report these analyses in the SOM.

<sup>11</sup> There are no significant differences between the 435 participants who accepted the Wave 2 study and the 80 participants who did not. These analyses are also reported in the SOM.

who identified with the same political party and held the same party-typical view as themselves on the issue of abortion access in the United States. Then, we randomly assigned predictors to either a “moderate” partner or an “extreme” partner condition and gave them information about their partner’s attitudes, which varied whether their ingroup partner either agreed with the typical view of the party between 50-85 on the agreement scale (which we refer to as the “moderate partner” condition) or 95-100 (which we refer to as the “extreme partner” condition). We informed predictors that their ingroup partner would learn the following information about them: (1) which political party they support, (2) that they were assigned to write a persuasive message in favor of the opposing party’s view on abortion access, and (3) that this writing task led them to report that their agreement with their party’s view on abortion access decreased. We then asked predictors to predict how their partner in the react condition would respond upon learning about their dissenting belief change.

We used a nine-item scale to measure anticipated social sanctions ( $\alpha = .97$ ), which consisted of the same five-item measure of social rejection from Study 1 and four additional items to measure reputation damage (Levine & Schweitzer, 2015). We included these new items to examine a broader breadth of interpersonal consequences, which asked participants how another ingroup member would react across the following dimensions after learning that their agreement with the typical view of their party decreased, “After the person you have been paired with learns that your agreement with the statement “abortion access in this country should be **protected/restricted**” **DECREASED**, how much do you think this person will feel inclined to...

(1) exclude you, (2) ignore your input, (3) reject you, (4) disrespect you, (5) criticize you (6) dislike you, (7) be upset with you, (8) lose respect for you, (9) lose trust in you?” (emphasis original; order randomized). All items were answered on a Likert-scale (1 = *Not at all*; 4 = *A*

*moderate amount*; 7 = *Very much*). We found that all nine of these items formed a single factor and showed high reliability ( $\alpha = .97$ ). Thus, we combined all items into a single composite measure, referred to henceforth as a measure of “social sanctions”, which we also used in Studies 3 and 4.

Participants who were randomly assigned to the react role condition were categorized as either “extreme” (those who responded between 95-100) or “moderate” (those who responded between 50-85) based on their Wave 1 responses to the 0-100 agreement measure. This participant-level quasi-independent variable served as the “extremity” condition factor in the react condition. We told reactors the exact information that we informed predictors we would tell them. Specifically, we told reactors that they were paired with another participant taking the study who identified with the same political party and agreed with the same party-typical view on the issue of abortion as themselves. We then told reactors that this person was assigned to write a persuasive message in favor of the opposing party’s point of view on the issue of abortion as part of the study, which caused their agreement with the typical view of the party to decrease. Reactors then reported social sanctions towards their partner on the same nine-item composite measures as the predictors (with modified wording so that reactions were captured instead of predictions).

We also included a behavioral measure of economic punishment in the form of a dictator game. We told participants that they would be randomly assigned to the distributor (dictator) versus receiver role for a financial bonus; however, in reality, reactors were always assigned to the distributor role. This paradigm created a financial incentive for reactors to punish or reward the dissenting ingroup member they were evaluating, wherein any amount they took from their partner represented a gain for themselves. For this item, predictors estimated how much of a

\$1.00 bonus their partner would give them versus keep for themselves and reactors distributed the bonus between themselves and their partner.

Lastly, after predictors completed all other dependent measures, they responded to a single item measuring how likely they would be to self-censor if their beliefs on abortion actually changed in the way they were asked to consider in the study. Specifically, predictors were asked, “If your agreement with the statement [piped text showing the typical view of their party] **actually DECREASED**, would you tell that to the person you’ve been paired with in this study if the topic came up in conversation?” (1 = *I definitely would NOT tell them*; 7 = *I definitely WOULD tell them*; emphasis original).

## **Results**

**Manipulation Check.** First, to examine whether “extreme” and “moderate” condition participants in our sample behaved as we anticipated, we conducted a series of independent samples t-tests to explore the effects of extremity on anticipated and actual social sanctions within the predict and react role conditions respectively. Predictors who were paired with an “extreme” ingroup member reported significantly higher anticipated social sanction scores ( $M = 4.31$ ;  $SD = 1.38$ ) compared to participants in the predict condition who were paired with a “moderate” partner ( $M = 3.93$ ;  $SD = 1.32$ ),  $t(197) = 2.00$ ,  $p = .047$ ,  $d = .28$ . Next, independent samples t-tests comparing extreme vs moderate reactors showed that “extreme” reactors reported higher social sanction scores ( $M = 2.88$ ;  $SD = 1.74$ ) than “moderate” reactors ( $M = 2.11$ ;  $SD = 1.26$ ),  $t(200) = 3.61$ ,  $p < .001$ ,  $d = .51$ .

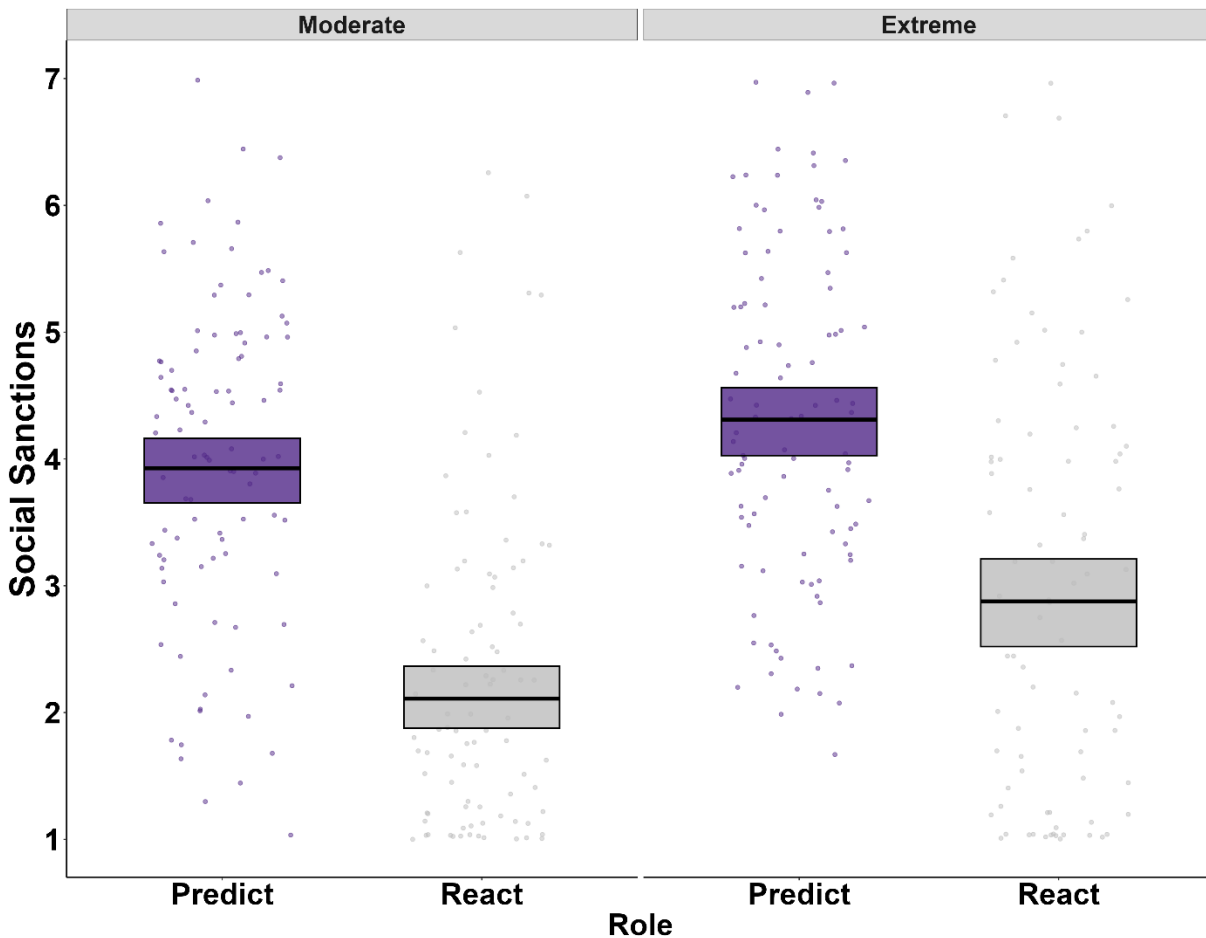
**Social Sanctions.** Next, we conducted a 2 (role: predict vs. react)  $\times$  2 (extremity: moderate vs. extreme) ANOVA on our composite dependent measure of social sanctions ( $\alpha =$

.97) to detect whether a perception gap was present between predictors and reactors, and, if so, whether the magnitude of the perception gap was greater for participants in the moderate condition versus extreme condition. In this analysis, the “extremity” variable took on the value of “partner extremity” for predictors and “participant extremity” for reactors. With this yoked design, we were able to match the predictions that predictors made about moderate and extreme ingroup members, respectively, with the social sanction scores that moderate and extreme reactors reported. This analysis provided further support for the perception gap, revealing a main effect of role wherein predictors ( $M = 4.12$ ;  $SD = 1.36$ ) significantly over-estimated sanctions relative to reactors’ reports ( $M = 2.47$ ;  $SD = 1.55$ ),  $F(1, 397) = 133.22$ ,  $p < .001$ ,  $\eta^2 = 0.25$ . We also observed a main effect of extremity,  $F(1, 397) = 16.18$ ,  $p < .001$ ,  $\eta^2 = 0.04$ ; however, the interaction of role  $\times$  extremity was not statistically significant,  $F(1, 397) = 1.81$ ,  $p = .18$ ,  $\eta^2 = 0.01$ . The lack of interaction in this model suggests that the effect size of role (i.e., the perception gap) was of comparable magnitude in the moderate and extreme conditions.

Next, we unpacked the simple effects of this model, testing whether the perception gap was present at each level of extremity condition (Figure 2). We found that predictors who were assigned to the extreme partner condition reported significantly higher social sanction scores ( $M = 4.31$ ,  $SD = 1.38$ ) compared to extreme reactors ( $M = 2.88$ ,  $SD = 1.74$ ),  $t(397) = 6.99$ ,  $p < .001$ ,  $d = 0.92$ . We found a similar pattern of results in the moderate partner condition, wherein predictors who were told they were paired with a moderate partner reported significantly higher social sanction scores ( $M = 3.93$ ,  $SD = 1.32$ ) than attitudinally moderate reactors ( $M = 2.11$ ,  $SD = 1.26$ ),  $t(397) = 9.11$ ,  $p < .001$ ,  $d = 1.41$ . These findings show that the perception gap was present at both levels of extremity; however, the effect size was slightly larger in the “moderate” condition compared to the “extreme” condition. This may have been because the range for



“moderate” attitudes (50-85) was relatively broader than the range for “extreme” attitudes (95-100). Thus, predictors in the “moderate” condition may have inferred their partners’ attitudes to be towards the upper end of this range, when in reality reactors' attitudes were distributed throughout this range.



**Figure 2.** Predictors report higher social sanction scores than reactors across extremity conditions. Note. The data in this plot are slightly jittered to avoid overlap. The height of the boxes represents the 95% confidence interval.

**Behavioral Punishment.** We conducted a 2 (role: predict vs. react)  $\times$  2 (extremity: moderate vs. extreme) ANOVA to examine whether predictors in our sample overestimated how

much reactors would punish them for dissenting belief change, and, if so, whether the magnitude of this overestimation was greater based on extremity condition. This analysis revealed a significant main effect of role,  $F(1, 397) = 14.10, p < .001, \eta^2 = 0.034$ , indicating that predictors anticipated that their react condition partner would keep significantly more of the bonus money than reactors actually did. The main effect of extremity was not significant,  $F(1, 397) = 1.61, p = .205, \eta^2 = 0.004$ , suggesting no overall difference between the moderate and extreme conditions. The interaction of role  $\times$  extremity was also not statistically significant,  $F(1, 397) = 0.09, p = .760, \eta^2 < 0.001$ , suggesting that the difference in the effect size of role was not significantly different between moderate and extreme partner conditions.

***Self-Censorship.*** Participants in the predict condition were asked whether they would share information about their belief change with their partner or self-censor if they actually updated their beliefs on the topic. This item was reverse-coded for a more intuitive interpretation in our analysis, with higher scores representing a greater likelihood of self-censorship. We conducted Pearson's correlational tests to examine the relationship between the likelihood of self-censorship, anticipated social sanctions, and anticipated economic punishment in the dictator game. These analyses revealed that predictors reported a greater likelihood of self-censorship when they anticipated harsher social sanctions ( $r = .22, p = .002$ ) and greater economic punishment in the dictator game ( $r = .26, p < .001$ ). These findings provide further in support of the relationship between fear of social sanctions and self-censorship observed in Pilot Study 1.

### ***Discussion***

Providing predictors with accurate information about their partner's attitudes did not close the perception gap. Instead, replicating Study 1, predictors in Study 2 significantly overestimated how much ingroup reactors would socially sanction them for dissenting belief

change in both extremity conditions. Predictors also overestimated how much they would be punished in the dictator game, providing evidence that the perception gap extends to contexts that involve tangible consequences. Taken together, these findings suggest that the overestimation of social sanctions is not driven by an overestimation of ingroup extremity but by other psychological factors.

### **Study 3: Paired Dyads & Behavioral Measures**

Taken together, Studies 1 and 2 provide further evidence in support of a perception gap between predicted and actual reactions to dissenting belief change. Although Study 2 suggests that perceptions of extremity do not drive the perception gap, there may be other stereotypes people bring to mind when trying to anticipate the reactions of someone they have not met. To explore whether this effect is present among ingroup members who have met and gotten to know each other, in Study 3 we created dyads of predictors and reactors who had a conversation and then completed our experimental paradigm. This relational context offers several methodological advantages and more closely approximates many real-life situations in which individuals grapple with whether or not to disclose dissenting belief change.

#### **Methods**

##### ***Participants***

We recruited participants from a University research subject pool of Amazon Mechanical Turk workers. We determined that we would need to recruit a sample size of at least 300 participants to detect a moderate effect size (*Cohen's d* = .5) at 90% power. Given that Study 3 occurred in two distinct phases and involved live participant pairing, we over-recruited heavily to ensure that our final sample was sufficient after accounting for attrition and excluding

participants who were not paired with a partner in the survey due to technical issues. We aimed to recruit 1,000 participants (500 Democrats, 500 Republicans) to complete Part 1 because we aimed to have a final sample in Part 2 of 500 (250 Democrats, 250 Republicans). We ended up with a final sample of  $N = 885$  in Part 1 due to time constraints.

The following day, participants were invited via email to participate in Part 2 of the study if they reported at least one political view that was aligned with the typical view of their political party on the topics of abortion, gun control, and immigration in the United States. Almost all participants ( $N = 817$ ; 92%) met this criterion. Among these, 278 participants (190 Democrats, 88 Republicans; 154 females, 124 males;  $M_{age} = 45.90$ ,  $SD_{age} = 12.04$ )<sup>12</sup> completed all study procedures, passed all attention checks, and were successfully paired with another participant according to our pre-registered pairing criterion (described below).

### ***Procedure***

*Part 1.* Participants completed a 2-minute political attitudes survey. They responded to political demographic and opinion questions, indicating their political orientation (Republican vs. Democrat); attitudes on the issues of abortion, immigration, and gun control (binary choice: liberal vs conservative stance, similar to Study 1); and how strongly they agreed with the selected stance on each of the topic (1 = *Strongly Disagree*; 7 = *Strongly Agree*). Participants also responded to basic demographic questions (age, gender, ethnicity).

*Part 2.* Once participants accepted the invitation to participate in Part 2, they were randomly assigned to either a predict condition or a react condition. Following a similar

---

<sup>12</sup> Due to a technical difficulty, a portion of participants who completed Part 2 were not successfully paired with another participant based on our criterion. These participants are not included in the analyses reported in the main text. Auxiliary analyses including all participants—even those for whom the randomization system failed—yielded identical conclusions for all hypothesis tests and are reported in the SOM.

procedure as Study 1, we randomly assigned participants to consider one of the political topics from Part 1 upon which they reported that they agree with the typical view of their party as the focal topic in Part 2 of the study.

Upon entering the survey, we told participants that the purpose of this study was to examine how well people perform on cooperative tasks when working with another person who holds either similar or different political beliefs. To that end, we told participants that they could be paired with an ingroup member or an outgroup member; however, in reality, all participants were paired with ingroup members who shared the same view on the randomly selected political topic.

We then informed participants that the study would proceed in the following order: First, they would have a brief conversation with their partner. Next, they would complete a brief survey (which contained our key dependent measures) on their own. Then, they would rejoin their partner to complete the cooperative task where they would have a chance to win a financial bonus based on their joint performance. In reality, the cooperative task did not exist. After participants completed all survey measures, we told them that they were disconnected from their partner due to a technical mishap that was of no fault of their own.

We used the pretense of a cooperative task because we wanted to create the anticipation of future interaction between participants. We expected that this would increase the stakes for cooperation and coordination between participants, thereby making our measures of anticipated and actual social sanctions both more consequential and realistic. The cover story also allowed us to include behavioral measures related to the task.

After reading about the study procedure, participants entered a chatroom using Chatplat—an online in-survey platform where participants can have a live chat conversation with

another participant in the study ([www.chatplat.com](http://www.chatplat.com); Brooks & Schweitzer, 2011)—and were paired with another participant. Using survey logic, participants were randomly paired with another participant who met the following criterion: (1) reported that they support the same political party, (2) was randomly assigned to the same political topic condition, upon which they also held the typical view of the party, and (3) was randomly assigned to the opposite role condition.

This pairing procedure allowed us to create dyads of predictor-reactor ingroup members who both held the party's typical view on the randomly assigned topic. As soon as another participant entered the chatroom who met this pairing criterion, the two participants were paired. Participants were informed, veridically, that their partner was another participant taking the survey who identified with the same political party and held the same view on the randomly selected topic as the participant (all study materials are reported in the SOM under Study 3 Additional Information). Participants completed two comprehension checks to ensure this information was understood before the chat began.

The creation of paired dyads in Study 3 offers three methodological advantages. First, this design allows us to establish a “ground truth” measure of predictor accuracy by comparing each predictor's estimate against the specific reactor who evaluated them. Second, this approach provides a more conservative test of the perception gap observed in previous studies, as participants may be better at taking the perspective of someone they have interacted with and are familiar with compared to a stranger (Stinson & Ickes, 1992; Thomas & Fletcher, 2003). Third, Study 3 places participants in an ostensible cooperative context where each participant's financial outcomes depend on their collaborative success with their partner. Under this pretense, we included two behavioral measures of social sanctions (dictator game, partner choice for future

interaction) in Study 3 to further examine whether the perception gap extends to behavioral consequences when predictors and reactors have interacted.

Once dyads were formed, participants were given five minutes to chat with their partner and were instructed to follow a simplified version of the “fast-friends” procedure (Aron et al., 1997) in which participants took turns asking each other get-to-know-you style questions<sup>13</sup>. When the conversation concluded, we told participants that they would soon reconnect with their partner to complete the cooperative task. On the next page, we gave participants the following message, “Before you are reconnected with your partner, there are a few things you should know about this experiment... PLEASE READ THE FOLLOWING INSTRUCTIONS CAREFULLY” (emphasis original). Participants were then given specific instructions based on which role condition they were assigned to.

We told predictors that at this point in the study, the research team was informing their partner that between Part 1 and Part 2 of the study, the predictor was assigned to watch a nonpartisan video from ProCon.org (which is a real, non-partisan informational website) that described both sides of the debate on the randomly assigned political topic. Although predictors never actually watched this video, we told them that their partner would learn that watching the ProCon video caused them to change their position on the randomly assigned topic. Specifically, participants were shown the same seven-point Likert-scale item upon which they reported their agreement with their party’s stance on the randomly assigned topic during Part 1 (1 = *strongly disagree*; 7 = *strongly agree*) and we told them that their partner would learn that their agreement moved one point to the left on the scale after watching the video. We showed predictors

---

<sup>13</sup> We chose this prompt because we wanted participants to get to know each other, but we did not want them to drift into discussing political topics, which may have reduced the realism of the ensuing experimental procedure.

screenshots of the information that their partner in the react condition would receive (which was true). These instructions implied that the predictor still agreed with the typical view of the party on the issue, but their agreement with this position decreased by the smallest measurable amount.

We gave reactors the exact information that we told predictors that we would tell their partner. Specifically, we told reactors that their partner was assigned to watch a ProCon.org video on the randomly selected political topic between Part 1 and Part 2 of the study. We told reactors that their partner still agreed with the typical view of the party, but watching this video led the participant to decrease their agreement with this position by one point to the left on the seven-point Likert item compared to how they responded to the same item in Part 1.

All participants were asked to confirm that they read and understood the instructions before responding to our dependent measures in the next part of the survey.

### *Measures*

First, all participants responded to the same 9-item composite measure of negative social consequences from Study 2. Next, participants responded to a free response item asking them to describe how they think their partner felt and why (predict condition) or how they felt and why (react condition) after learning the information about belief change.

Next, participants responded to two behavioral measures of social sanctions. Participants first read a basic description of the cooperative task they believed they were about to complete, which was described as navigating a virtual maze as a team. We then told participants that their bonus payout would depend upon their ability to successfully cooperate with their partner to complete this task. In other words, participants were led to believe that their financial outcomes were linked.



After reading the task description, participants responded to a binary-choice behavioral measure of partner choice. Specifically, reactors chose whether they wanted to work on the cooperative task with a different partner or maintain the same partner. We told predictors that their partner would have this choice and asked them to predict whether their partner would choose to work with someone else or with them on the task.

Next, participants completed an adapted version of a dictator game. Participants were told that if they successfully completed the task, they would be awarded a \$0.10 bonus that one person in the dyad would be randomly assigned to distribute between them. In reality, reactors were always assigned the role of bonus distributor and reactors chose how to distribute the \$0.10 bonus between themselves and their partner. Predictors were asked to predict how their partner would distribute the bonus.

After participants responded to all survey measures, they reached the end of the study. We told participants that they were disconnected from their partner due to a technical issue that was no fault of their own and that, consequently, they would not be able to attempt the cooperative task. This information came at the very end of the study and did not influence participant survey responses, and all participants were awarded a \$0.10 bonus. To conclude, we gave participants a debriefing form that described the use of deception in the study and provided the rationale behind the study procedure.

## ***Results***

Were predictors estimates accurate? Findings from Study 3 provided further evidence in support of the perception gap observed in Study 1 and Study 2. Specifically, predictors expected that their partner would socially sanction them ( $M = 3.14$ ;  $SD = 1.14$ ) for dissenting belief

change significantly more than reactors reported ( $M = 2.02$ ;  $SD = 0.84$ ),  $t(276) = 9.32$ ,  $p < .001$ ,  $d = 1.12$ . In total, 110 predictors (79%) overestimated social sanctions relative to what their react condition partner reported (7 predictors were accurate; 22 underestimated). We note that the means for social sanctions in both predict and react conditions are slightly lower than we have observed in previous studies; however, the effect size of the perception gap is similar. This pattern may suggest the possibility that interpersonal contact can buffer absolute expectations and actual social sanctions, even as the effect size of overestimation remains the same.

Next, we analyzed predictor and reactor responses to the free response items, which asked them to describe how they thought their partner felt (predictors) or how they felt (reactors) after receiving the information about the dissenting belief change. We assembled a team of four hypothesis-blind research assistants who independently coded participant responses ( $ICC = .86$ ) based on the expected and actual effects of this information on interpersonal evaluations ranging from -3 (very negative) to +3 (very positive). Zero in this coding scheme represented “neutral / no effect” on the interpersonal evaluation. Coders rated a five for any response that did not answer the question and all rows that were rated a five were excluded from analysis, yielding a sub-sample of 219 participants whose responses to these questions were coded. Participant free-response data were consistent with what we observed on survey measures: Participants in the predict condition anticipated significantly worse interpersonal outcomes ( $M = -0.83$ ,  $SD = 0.97$ ) than participants in the react condition reported ( $M = 0.00$ ,  $SD = 1.04$ ),  $t(215) = 6.06$ ,  $p < .001$ ,  $d = .83$  (Figure 3; see Study 3 Additional Information in the SOM for the full coding scheme and instructions).

Lastly, we observed a similar pattern of results on the behavioral measures. On average, predictors overestimated how often their partner would choose to work with someone else on the

cooperative task (Predict: 18.71%; React: 7.91%),  $X^2(1, 278) = 6.11, p = .013, V = .15$ , and expected their partner to keep significantly more of the bonus ( $M_{predict} = 5.72$  cents;  $SD = 1.58$ ) than their reactor partners kept ( $M_{react} = 5.18$  cents;  $SD = 1.43$ ),  $t(276) = 2.99, p = .003, d = .36$ .

These findings help to rule out the possibility that reactors responded to survey item measures in a socially desirable way (e.g., trying to appear open-minded and accepting) because of the financially incentivized stakes of their behaviors. Specifically, reactors should have been inclined to penalize their predict condition partner given that (a) any amount of money they gave to their partner in the dictator game would have lowered their own bonus, and (b) their ability to win the bonus hinged upon their ability to cooperate with their partner. Instead, we observed that reactor behaviors were congruent with their attitudes, distributing the bonus more generously and choosing to reject their partner far less often than predictors expected. Moreover, participant qualitative responses were convergent with survey and behavioral findings, providing a comprehensive view of participants' anticipated and actual reactions.



**Figure 3.** *Predictors' free response text data reveal that they expected their partner would be more upset than their partners actually were. Note.* The plot shows the density distribution of coded scores for predictor and reactor responses. Density reflects the relative concentration of scores at different points on the scale, with higher peaks indicating more frequent or typical responses. The vertical dashed lines represent the mean coded sentiment for each condition.

### ***Discussion***

Study 3 replicated the pattern of overestimated social sanctions observed in Studies 2 and 3 across survey, behavioral, and qualitative measures. Using paired dyads, Study 3 provided definitive evidence that predictors' estimates about their partner's evaluative and behavioral responses were indeed inaccurate.

### **Study 4: Actual Belief Change**

Studies 2 and 3 documented that group members overestimate the social costs of dissenting belief change across varying relationship contexts and degrees of belief change; however, we have yet to document the overestimation of social sanctions among predictors who have actually changed their beliefs. Given that people are motivated to see themselves as rational and consistent (Pronin et al., 2002; Ross & Ward, 1996); it is plausible that asking predictors to imagine changing their minds in this way may have seemed absurd, and therefore, deserving of social sanctions. We addressed this limitation in Study 4, which occurred in two phases several months apart.

#### **Phase 1 Method**

##### *Participants*

Our sample size for Phase 1 of Study 4 was determined by challenges in recruiting a specific sub-sample of a population during Phase 1 (i.e., participants whose beliefs changed after writing a counter-attitudinal message). For Phase 1, we recruited a total of 1,524 participants from Amazon's Mechanical Turk and Prolific to take part in a research study<sup>14</sup>; however, only a small subset of participants in the predict condition reported dissenting belief change. To address this limitation and achieve a sufficient sample size, we combined two samples of participants who completed a nearly identical study procedure. Combining these samples for Phase 1 created a larger pool to re-recruit from for Phase 2 of the study. Thus, our final sample for Phase 1

---

<sup>14</sup> Due to challenges in collecting a large sample of participants who reported dissenting belief change following the counter-attitudinal writing task, participants in Phase 1 of Study 4 were collected from three separate studies with a similar experimental procedure. The methods and findings for each of the three studies are nearly identical and are reported in full in the SOM.

consisted of  $N = 494$  participants (147 predictors, 347 reactors; 262 Democrats, 232 Republicans; 270 females, 225 males, 5 self-identify;  $M_{age} = 42.53$ ,  $SD_{age} = 13.05$ ).

### ***Procedure***

The goal of Phase 1 was to test whether predictors who have experienced dissenting belief change on a central political topic overestimate ingroup social sanctions. This study employed a similar design as previous studies wherein participants were randomly assigned to either a predict or a react condition. The beginning of Phase 1 was similar to previous studies wherein participants reported their partisan identity and their attitudes on central political topics (abortion, gun control, immigration). After being assigned to consider a single political topic upon which participants reported a view that was consistent with the typical view of their party, predictors were assigned to complete a counter-attitudinal writing task in which they wrote a persuasive message in favor of the opposing party's view on that topic. We employed this procedure because previous research has shown that it produces a relatively consistent effect of belief change outcomes in a similar context (e.g., Briñol et al., 2012; Carlsmith et al., 1966; Greenwald & Albert, 1968). Predictors reported how strongly they agreed with the typical view of their party both before (T1) and after (T2) completing the counter-attitudinal writing task using a 0-100 slider scale (0 = *Strongly Disagree*; 100 = *Strongly Agree*). We determined whether a predictor reported a dissenting shift in their beliefs by calculating a difference score between their T1 and T2 agreement measures. Any predictor who reported a score that was lower at T2 than it was at T1 was categorized as a dissenter. Dissenting predictors then reported how they expected another ingroup member would react if they learned about their completion of this task and its effects on their beliefs using the same nine-item composite measure of social sanctions from Study 2.

We told reactors that they would be randomly assigned to evaluate a participant from the predict condition. Reactors saw screenshots of information that was used to describe the counter-attitudinal writing task to predictors and were told that completing this task led to dissenting belief change away from the typical view of the party (see SOM Study 4 Additional Information for all study materials).

### ***Phase 1 Results***

To examine whether predictors overestimated the social costs of dissenting belief change, we conducted an independent samples t-test comparing social sanction scores between predict and react conditions. Consistent with findings from our previous studies, results revealed that predictors who reported dissenting belief change anticipated significantly harsher social sanctions ( $M = 3.61$ ;  $SD = 1.63$ ) compared to reactors' reports ( $M = 2.74$ ;  $SD = 1.63$ ),  $t(492) = 5.41$ ,  $p < .001$ ,  $d = 0.53$ .

### **Study 4 Phase 2**

Findings from Phase 1 replicate the main effect of overestimation observed in Studies 2 and 3, demonstrating that the effect persists with actual belief change. However, an interpretational challenge is that it is essentially impossible to assign people at random to change (or not change) their attitudes on a divisive moral issue, and it is possible that some unaccounted-for person-level variable led predictors to both change their beliefs and overestimate social sanctions. Given that only a small subset of participants in Phase 1 reported dissenting belief change, we recognize that this limits our ability to draw inferences from these data. To push toward causal evidence despite the stickiness of strong moral attitudes, Phase 2 randomly

assigned participants either to a condition in which their dissenting belief change was made salient or to a control condition.

In Phase 2 of Study 4, dissenting predictors from Phase 1 were invited to participate in a study after a delay period of approximately two months, which we expected would be sufficient time for most predictors to forget about their participation in Phase 1. During Phase 2, participants were randomly assigned to either receive a tailored reminder about the task they completed and their dissenting belief change during Phase 1 or to receive a vague, non-specific reminder about their earlier participation with no specific mention of their dissenting belief change (see the SOM Study 4 Additional Information for a full description of the reminder in each condition). All participants then predicted how an ingroup member would react if they learned about their participation in Phase 1.

This design allowed us to accomplish two goals central to the research program. First, it allowed us to test the causal relationship between (the salience of) dissenting belief change and the heightened expectation for social sanctions. Second, it afforded tests of a competing explanation for the misperception documented in previous studies: that participants have been exhibiting a generalized form of undersociality bias, in line with previous findings showing that people tend to overestimate negativity during social interactions (see Epley et al., 2022, for review). From this perspective, the perceptual gap observed in our studies may simply be the manifestation of a ubiquitous bias in social judgment wherein people underestimate how much they will be liked by others, which would be present regardless of belief change. If this is indeed the case, then participants in both conditions in Phase 2 should anticipate social sanctions similarly.

## **Phase 2 Method**



### *Participants*

The sample size for Phase 2 was determined by participant attrition over a delay period of approximately two months between Phase 1 and Phase 2. We invited all predictors from Phase 1 who reported dissenting belief change to participate in Phase 2 of the study. Among these, 63% completed Phase 2, leaving us with a final sample size of  $N = 93$  for Phase 2.

### *Procedure*

After a delay period of two months, participants entered the study and were randomly assigned to either have their belief change from Phase 1 made salient (“salient condition”) versus not made salient (“non-salient condition”) The instructions participants saw are shown below. Participants in the control condition only saw the first paragraph. Participants in the salient condition received the full message.

Thank you for participating in this study. You have been invited to participate in this study because earlier this fall you participated in one of our research studies and we would like to ask you some follow up questions.

[salient condition only from here on] You participated in one of our research studies on November 30th. At the beginning of that study, you indicated that you support the following view on the topic of abortion access in the United States: “abortion access in this country should be **[protected/restricted]**”.

Later on in that study, you were asked to write a persuasive message on the topic of abortion access in the United States in favor of the following opposing political stance: “abortion access in this country should be **[protected/restricted]**”.

After you finished writing a message in favor of the opposite view on this topic, you reported that your agreement with the statement, “abortion access in this country should be [protected/restricted]” **DECREASED**. That is, your agreement with the majority position of the [ingroup] party on the topic of abortion access in the United States decreased after writing a persuasive message in favor of the majority position of the [outgroup] party.

Following these condition-specific instructions, all participants were then asked to predict how an ingroup member would react upon learning about their participation in the previous study using the same nine-item composite measure of social sanctions from previous studies. To conclude the study, participants responded to a manipulation-check question which asked participants to recall specific details about their belief change during Phase 1 as a test of whether these details were salient.

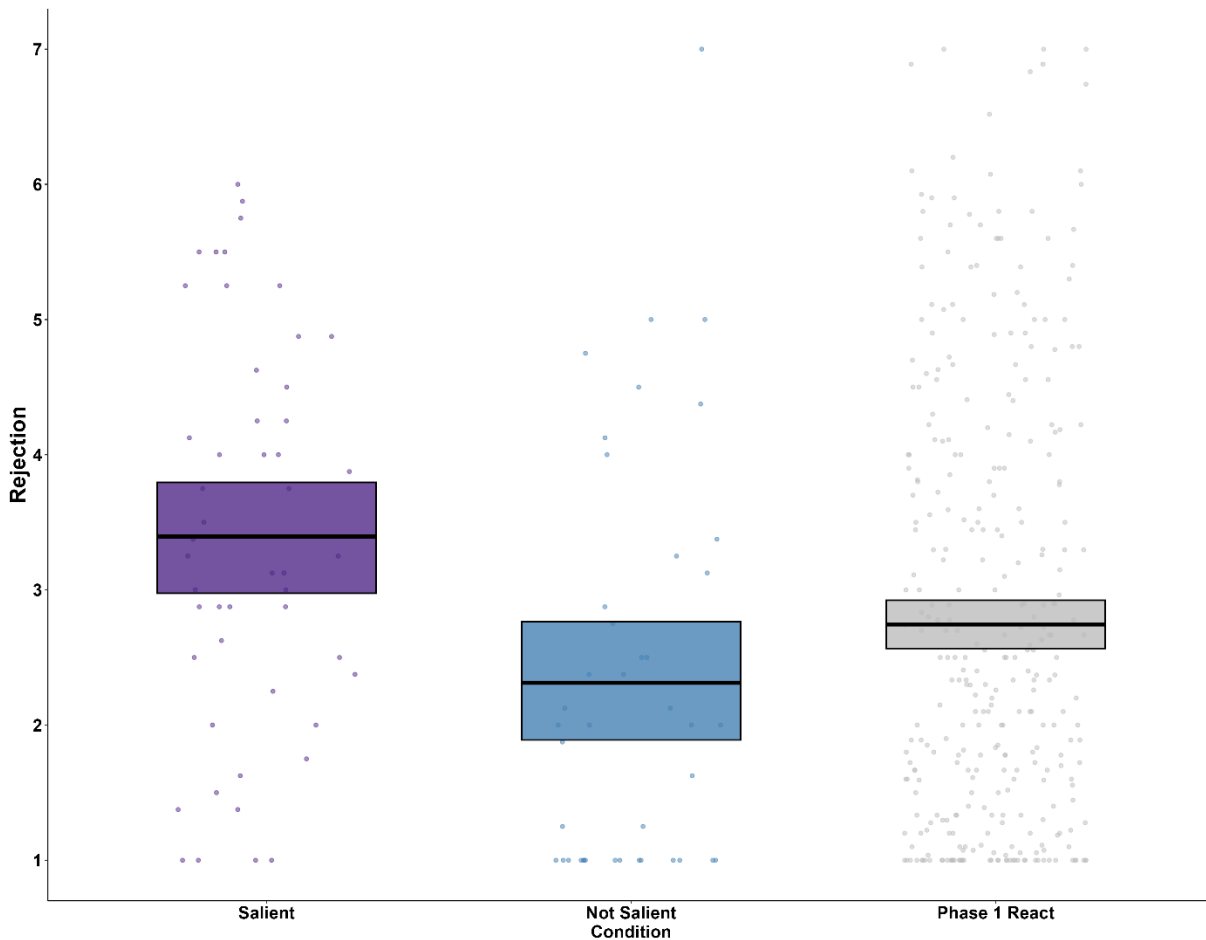
### ***Phase 2 Results***

***Manipulation Check.*** In response to the manipulation check item, 74% of participants in the salient condition accurately recalled details about their belief change, whereas only 22% of participants in the non-salient condition were accurate. These findings confirm that dissenting belief change was more salient for participants in the salient condition compared to the non-salient condition.

Next, we tested whether expectations for social sanctions were different between the salient and non-salient conditions. Participants in the salient condition ( $M = 3.40$ ;  $SD = 1.45$ ) reported that they anticipated significantly more social sanctions than participants in the non-salient condition ( $M = 2.31$ ;  $SD = 1.48$ ),  $t(90) = -3.53$ ,  $p < .001$ ,  $d = .74$ . Next, we conducted a

one-way ANOVA to compare predictors' social sanction estimates in each Phase 2 condition against reactor reports of social sanctions towards dissenting ingroup members from Phase 1.

The main effect of this test was significant,  $F(2, 436) = 5.60, p = .004, \eta^2 = .025$ . Post-hoc comparisons using Tukey's HSD test indicated that predictors in the Phase 2 salient condition ( $M = 3.40; SD = 1.45$ ) overestimated social sanctions compared to reactor scores from Phase 1 ( $M = 2.74; SD = 1.63$ ),  $p = .020, d = .41$ . In contrast, participants in the Phase 2 non-salient condition did not overestimate reactor social sanctions ( $M = 2.31; SD = 1.48$ ),  $p = .224, d = .27$  (Figure 4).



**Figure 4.** *Dissenting predictors in Phase 2 whose belief change was made salient overestimated social sanctions. Dissenting predictors whose belief change was not made salient did not. Note.*

The data in this plot are slightly jittered to avoid overlap. The height of the boxes represents the 95% confidence interval.

### *Discussion*

Study 4 extends the present research in two ways. First, it replicates the perception gap documented in Studies 2 and 3—which relied on hypothetical scenarios—using a sample of participants who reported dissenting belief change on a central political topic during the study. Second, Study 4 demonstrates the causal effect of dissenting belief change on anticipated social sanctions: Among predictors who reported dissenting belief change, only those who received a reminder about this change overestimated social sanctions in Phase 2.

These findings have several implications. First, they address a major limitation of Studies 2 and 3 by showing that the perception gap replicates with actual belief change. Second, they provide evidence of a causal relationship between (the salience of) dissenting belief change and anticipated sanctions. Third, they help to rule out one viable alternative explanation for this perception gap: Undersociality bias. Previous studies have shown that people robustly and consistently underestimate how positive social interactions will be (see Epley et al., 2022, for review). If a general under-sociality bias was driving misperceptions of social sanctions, then we should expect that participants from Phase 2 in both the salient and non-salient conditions would similarly overestimate social sanctions. In contrast, by showing that participants do not overestimate social sanctions when their dissenting belief change is not salient, we provide evidence in support of the unique causal role of the salience of dissenting belief change in driving the overestimation of social sanctions.

Finally, the finding that predictors only overestimate when their belief change is made salient may suggest that psychological distance—how psychologically “close” versus “distant” one’s change in belief feels—influences their estimates of social sanctions. Simply imagining the potential of future rejection can be a stress-inducing experience (Gunther et al., 2010), which can heighten egocentric bias and impair the ability to reason about others’ internal states (Todd et al., 2015). Consequently, predictors—sensitive to the potentially debilitating effects of social rejection—fail to predict how others would react to dissenting belief change. Therefore, one route to mitigate the effects of egocentric bias and improve social judgment may be to foster psychological distance from the risk of rejection.

### **Study 5: Correcting Egocentric Bias to Correct the Misperception**

Across Studies 2-4, we have shown that partisans overestimate the social costs of dissenting belief change. What causes these misperceptions? We hypothesize that predictors exaggerated expectations are rooted in egocentric bias—which impairs perspective taking—and that attenuating egocentric bias will help to narrow this perception gap. Research on psychological distancing suggests that adopting a third-party perspective reduces egocentric bias and improves perspective taking (Golubickis et al., 2016; Grossmann et al., 2016; Kross & Grossmann, 2012; Staudinger & Glück, 2011; White et al., 2019). Thus, we hypothesize that the perception gap will close when predictors estimate social rejection from a third-party perspective compared to when they make the same estimate from a first-person perspective<sup>15</sup>.

### **Method**

---

<sup>15</sup> In preparation for Study 5, we conducted a small pilot study ( $N = 159$ ) using a similar design as Study 5 that replicated the pattern of results observed in Study 5. This pilot is reported in the SOM as Pilot Study 5.

### ***Participants***

Study 5 used a 3-cell between-participants design. We recruited 700 participants to yield at least 200 participants per cell after making exclusions for failed attention and comprehension checks. This sample size was determined based on a priori power analysis that indicated 200 participants per cell would provide sufficient power to detect our smallest hypothesized effect size ( $f = .15$ ) at 90% power with alpha set to .05. Far fewer participants failed attention and comprehension checks than we anticipated, yielding a final sample of  $N = 693$  after making such exclusions (510 Democrats, 183 Republicans; 426 females, 260 males, 7 self-identify;  $M_{age} = 41.03$ ;  $SD_{age} = 13.60$ ).

### ***Procedure***

Similar to previous studies, Study 5 began with a brief set of political attitude questions wherein participants reported their partisan identity and their attitudes on three central political topics (abortion, gun control, and immigration). Using the same procedure as Study 1, we then randomly assigned participants to consider one of the topics upon which they reported a view that was consistent with the typical view of their political party as the focal topic in the study. We then randomly assigned participants to either a first-person condition, a third-party condition, or a react condition. Participants in the first-person condition were asked to predict how much they would be rejected by another ingroup member for adopting the opposing party's view on the randomly assigned topic, similar to the predict condition procedure in Study 1. Participants in the react condition reported rejection towards an ingroup member who adopted the opposing party's view on the randomly assigned topic, similar to the react condition procedure in Study 1. Participants in the third-party condition were asked to predict how an ingroup member would react if another ingroup member (named Sam) adopted the opposing party's view on the

randomly assigned topic. An example of the third-party condition for a Republican participant assigned to the gun control condition is shown below (emphasis original).

Another **Republican** participant taking this survey named Sam indicated that they believe that access to private gun ownership in this country should be **protected**.

In the next part of the survey, we are going to ask you questions about how you think **another Republican on Prolific** would react if they found out that Sam changed their mind on this issue and now believes that access to private gun ownership in this country should be **restricted**.

After reading condition-specific instructions, all participants responded to the five-item social rejection composite measure that was used in Study 1. At the end of the study, we showed participants the instructions they saw earlier in the study and asked a manipulation check question that has been used in previous studies to measure psychological distancing (White et al., 2019): “When you imagine this specific situation, to what extent do you feel like you are an observer of this situation (i.e., see the situation from the perspective of an outside observer) versus an immersed participant in the situation (i.e., see the situation through your own eyes as if you are in it)?” (1 = *I feel like an observer*; 7 = *I feel like an immersed participant*).

Participants also completed two attention checks and one comprehension check. The first attention check asked participants about the political topic they were asked to consider during the study. The second attention check asked participants about the party identity of the person they read about during the study. The comprehension check required participants to indicate which of two situations represented a distanced observer’s perspective (“Two people on the other side of the street get into an argument”) versus an immersed participant’s perspective (“You meet a

friend for lunch and discuss your personal life”). We included this comprehension check before our manipulation check to familiarize participants with the terms used in the scale points of the manipulation check question (“distanced” and “immersed”). Per our pre-registration, we excluded participants who failed either of the attention checks from the analysis. We excluded participants who failed the comprehension check from analyses of the manipulation check item.

We hypothesized that we would detect a significant main effect of condition (first-person vs. third-party vs. react) on the social rejection composite such that participant scores would be highest in the first-person condition, second highest in the third-party condition, and lowest in the react condition<sup>16</sup>.

## **Results**

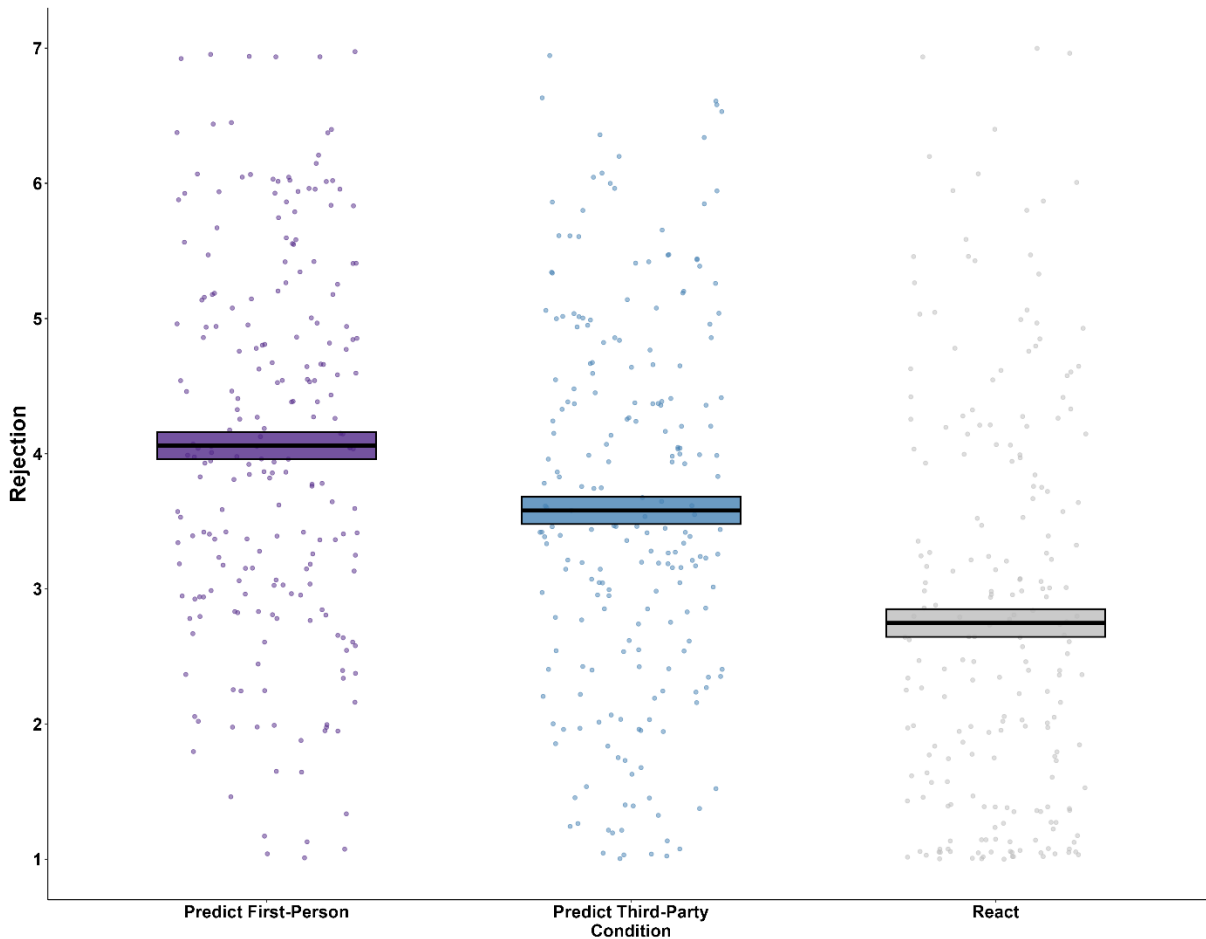
**Manipulation Check.** We conducted an ANOVA to determine the effect of between-participants condition on manipulation check scores after excluding participants who failed the comprehension check about psychological immersion versus distance. The results of the ANOVA indicated a significant effect of condition on psychological distance,  $F(2,630) = 78.87$ ,  $p < .001$ ,  $\eta^2 = 0.20$ , confirming that our experimental conditions affected participants’ feelings of psychological distance versus immersion. We conducted post-hoc comparisons using Tukey’s HSD test, which indicated that participants in the third-party condition ( $M = 2.34$ ,  $SD = 1.68$ ) reported significantly greater psychological distance compared to participants in the first-person condition ( $M = 4.59$ ,  $SD = 1.91$ ),  $t(630) = 12.80$ ,  $p < .001$ ,  $d = 1.25$ , and compared to participants in the react condition ( $M = 3.97$ ,  $SD = 2.07$ ),  $t(630) = 8.86$ ,  $p < .001$ ,  $d = .86$ .

---

<sup>16</sup> This study also included an exploratory within-subjects factor that produced null results. These methodological details, analyses, and results are reported in the SOM.



To explore our main hypothesis, we conducted an ANOVA to determine the effect of condition on social rejection composite scores. In support of our central hypothesis, we detected a significant effect of condition,  $F(2, 690) = 43.45, p < .001, \eta^2 = 0.11$ . Next, we conducted planned orthogonal contrasts to test our hypotheses that participants in the first-person condition would report higher scores than participants in the react and third-party conditions. In a replication of the primary misperception finding from the previous studies, the first planned contrast revealed that scores in the first-person condition ( $M = 4.06, SD = 1.52$ ) were significantly higher than scores in the react condition ( $M = 2.75, SD = 1.57$ ),  $t(690) = -3.35, p < .001, d = .85$ . The second planned contrast showed that scores in the first-person condition were significantly higher than scores in the third-party condition ( $M = 3.58, SD = 1.52$ ),  $t(690) = 5.82, p < .001, d = .32$ . The third planned contrast indicated that scores in the third-party condition were significantly higher than scores in the react condition,  $t(690) = -9.21, p < .001, d = .53$  (Figure 5).



**Figure 5.** *Participants are more accurate at predicting ingroup member reactions to dissent from a third-party perspective compared to a first-party perspective. Note.* The data in this plot are slightly jittered to avoid overlap. The height of the boxes represents the 95% confidence interval.

### ***Discussion***

Findings from Study 5 supported our hypothesis that participants would expect significantly harsher social consequences for dissenting belief change from a first-person perspective than from a third-party perspective. These findings highlight the link between a risk-averse orientation towards social rejection and egocentric bias, showing that participants

estimates are more accurate when adopting a psychologically distant perspective. We note, however, that the perception gap did not close entirely after adopting a third-party perspective. Research on perspective taking has shown that individuals reason about others' perspectives through a process of anchoring-and-adjustment, wherein they begin with their own egocentric perspective and then attempt to make adjustments to take on others' points of view (Epley et al., 2004). In this context, we suspect that participants in the third-party condition began with their own egocentric perspective (i.e., considering how much they themselves would be rejected) and then adjusted accordingly to estimate for a third-party.

### **General Discussion**

Across five studies and three pilot studies reported in the SOM, we consistently found that U.S. partisans overestimate the negative social reactions they would face from their political ingroup for dissenting belief change. The perception gap was large (weighted average effect size across studies 1-5:  $d = .84$ ) and robust across various contexts: it occurred for strangers (Studies 1, 2, 4, 5) and acquaintances (Study 3), for hypothetical (Studies 1, 2, 3, 5) and actual belief change (Study 4), and for complete (Studies 1, 5) and incremental shifts in beliefs (Studies 2, 3, 4). We found evidence for the perception gap across an array of measures, including survey items (Studies 1-5), behavioral observations (Studies 2, 3), and qualitative assessments (Study 3). These systematically miscalibrated expectations predicted communication behaviors; the more participants anticipated rejection, the more likely they were to self-censor dissenting belief change (Pilot Study 1, Study 3).

We demonstrated that the perception gap could be narrowed by reducing egocentric bias through a third-party perspective taking intervention. Specifically, when participants estimated ingroup reactions from a third-party perspective rather than from a first-person perspective, the

overestimation of rejection significantly decreased (Study 5). In addition, we empirically tested two alternative mechanisms that did not seem to explain the perception gap: that predictors overestimate reactor extremity (Study 2) and that predictors are simply pessimistic towards social interactions in general (Study 4).

In sum, we find persistent evidence of a perception gap wherein group members—limited by their own egocentric perspective—systematically overestimate social rejection from other group members for dissenting belief change. These findings have notable implications for social psychological theories of social misperceptions, self-censorship, and social conformity, as well as everyday social interactions.

## **Novelty and Theoretical Contributions**

### *Social Misperceptions*

First, we document a novel and robust social misperception among political ingroup members. Second, our research adds to the rich and growing literature on social misperceptions by showing that egocentric bias blunts perspective taking and skews predictions of social rejection. Evidence from other studies across numerous social domains converges on the idea that people tend to underestimate how well social interactions will go (i.e., “undersociality”; e.g., talking to strangers, Epley & Schroeder, 2014; sharing secrets, Kardas, Kumar, & Epley, 2023; discussing political disagreement, Wald et al., 2024). We believe that our findings—interpreted through the lens of humans’ inherent “better-safe-than-sorry” orientation towards rejection—help to explain part of the psychological process that produces pessimistic expectations about social interactions by showing that self-protective egocentric bias skews social judgment and creates misperceptions.

Our findings further contribute to this literature by bridging research on psychological distancing and social misperceptions. We show that third-party perspective taking promotes psychological distancing and reduces egocentric bias, thereby allowing for more accurate social judgments. This approach to correct misperceptions may prove effective in other contexts where social misperceptions act as a barrier to social interaction (Epley & Schroeder, 2014; Kardas et al., 2024; Wald et al., 2024).

Lastly, our work contributes to a growing body of research on group-level misperceptions in the realm of politics (Fernbach & Van Boven, 2022). Whereas previous studies have focused on misperceptions *between* political groups (Ruggeri et al., 2021; Lees & Cikara, 2020; Mernyk et al., 2022; c.f., Dias et al., 2024; Nyhan & Reifler, 2010), we document a robust and consequential misperception *within* political groups. The present work also adds to this literature by introducing a novel approach to correcting misperceptions that focuses on addressing their underlying psychological source (in this case, egocentric bias).

### *Self-Censorship & Social Conformity*

Our study of intragroup misperceptions also contributes to existing research on self-censorship and social conformity. We found that group members were increasingly likely to self-censor dissenting belief change insofar as they expected to be rejected for doing so (Pilot Study 1, Study 2). Although psychological theories of deviance and conformity have traditionally focused on social reactions to dissent (Jetten & Hornsey, 2014; Kruglanski & Webster, 1991), our work suggests that dissenters' expectations about social reactions they may elicit are also a powerful force of conformity. Taken together, these results imply that correcting misperceptions about ingroup rejection could potentially encourage dissent within groups.

*Benefits of Dissent*

Encouraging dissent produces numerous benefits for groups. By challenging the status quo, dissenters prevent “groupthink” and promote critical thinking, leading to higher-quality judgment and decisions (Duarte et al., 2015; Nemeth & Goncalo, 2011; Nemeth, 1995). Dissent can also encourage other group members to re-evaluate their own assumptions (Nemeth et al., 2001), which can prevent the group from shifting toward more polarized, extreme positions (Jetten & Hornsey, 2014; Nemeth & Goncalo, 2011). By correcting misperceptions, we propose that it may be possible to lower the psychological barriers to dissent, thereby encouraging the expression of diverse viewpoints and preventing polarization within groups.

**Limitations and Future Research**

Despite the robustness and consistency of the present findings across studies, several limitations are acknowledged. First, this research was conducted within the specific political and cultural context of the United States during a time of heightened political polarization. This context may limit the generalizability of our findings to other cultural or political settings.

Second, our studies primarily focused on exchanges between participants who were often strangers or had limited prior acquaintance. Although this allowed us to control for certain variables and focus on the psychological mechanisms at play, it may not fully capture the complexities of real-world interactions, particularly within close relationships. The stakes of social rejection are likely to be higher when it involves family members, friends, or colleagues, and the dynamics of self-censorship and social judgment may differ in these contexts.

Third, our studies were conducted in the context of dyadic exchanges, which may not fully capture the dynamics of larger group interactions, particularly in online environments. The

public nature of social media platforms, where a single post can instantly reach a large and diverse audience, may amplify both perceived and actual risks of rejection. Future research should examine how these dynamics unfold in online settings and larger in-person social interactions, where the potential for social sanctioning may differ significantly from the contexts studied here.

Finally, we identified that this perception gap is due, in part, to egocentric bias by showing that third-party perspective taking produces more accurate social judgments. Although the third-party perspective taking intervention reduced the size of the perception gap by 37.65% (from  $d = .85$  to  $d = .53$ ), it did not fully close the gap. Future research should examine whether stronger distancing interventions may close this gap, as well as other processes that may contribute to this effect.

## **Conclusion**

Our research sheds light on the complex dynamics of social judgment and social conformity within political groups, revealing a significant perception gap between expected and actual social reactions to dissenting belief change. By identifying egocentric bias as a key driver of this misperception, these studies offer new insights into the psychological processes that shape expectations about ingroup reactions to dissent and how such expectations, in turn, influence communication behaviors within groups. Our work underscores how a risk-averse orientation to social rejection creates tension between the need to belong and the importance of dissent within political groups. By addressing the psychological roots of social misperceptions—particularly within groups rather than between them—we can move toward a more inclusive and representative public discourse.

### References

- Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin*, *23*(4), 363–377.
- Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1.
- Ayduk, Ö., & Kross, E. (2010). From a distance: implications of spontaneous self-distancing for adaptive self-reflection. *Journal of Personality and Social Psychology*, *98*(5), 809.
- Bakker, B. N., Lelkes, Y., & Malka, A. (2020). Understanding partisan cue receptivity: Tests of predictions from the bounded rationality and expressive utility perspectives. *The Journal of Politics*, *82*(3), 1061–1077.
- Batteux, E., Ferguson, E., & Tunney, R. J. (2019). Do our risk preferences change when we make decisions for others? A meta-analysis of self-other differences in decisions involving risk. *PloS One*, *14*(5), e0216566.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529.
- Baumeister, R. F., & Tice, D. M. (1990). Anxiety and social exclusion. *Journal of Social and Clinical Psychology*, *9*(2), 165-.
- Beisswanger, A. H., Stone, E. R., Hupp, J. M., & Allgaier, L. (2003). Risk taking in relationships: Differences in deciding for oneself versus for a friend. *Basic and Applied Social Psychology*, *25*(2), 121–135.



- Brady, W. J., McLoughlin, K. L., Torres, M. P., Luo, K. F., Gendron, M., & Crockett, M. J. (2023). Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility. *Nature Human Behaviour*, 7(6), 917–927.
- Briñol, P., McCaslin, M. J., & Petty, R. E. (2012). Self-generated persuasion: effects of the target and direction of arguments. *Journal of Personality and Social Psychology*, 102(5), 925.
- Brooks, A. W., & Schweitzer, M. E. (2011). Can Nervous Nelly negotiate? How anxiety causes negotiators to make low first offers, exit early, and earn less profit. *Organizational Behavior and Human Decision Processes*, 115(1), 43–54.
- Brown, J. R., & Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour*, 5(8), 998–1008.
- Carlsmith, J. M., Collins, B. E., & Helmreich, R. L. (1966). Studies in forced compliance: The effect of pressure for compliance on attitude change produced by face-to-face role playing and anonymous essay writing. *Journal of Personality and Social Psychology*, 4(1), 1.
- Cavazza, N., Pagliaro, S., & Guidetti, M. (2014). Antecedents of concern for personal reputation: The role of group entitativity and fear of social exclusion. *Basic and Applied Social Psychology*, 36(4), 365–376.
- Chopik, W., Götschi, K., Carrillo, A., Weidmann, R., & Potter, J. (2024). Changes in need for uniqueness from 2000 until 2020. *Collabra: Psychology*, 10(1).
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55(1), 591–621.
- Connors, E. C. (2020). The social dimension of political values. *Political Behavior*, 42(3), 961–

982.

Correll, J., & Park, B. (2005). A model of the ingroup as a social resource. *Personality and Social Psychology Review*, 9(4), 341–359.

Dias, N. C., Aarslew, L. F., Frederiksen, K. V. S., Lelkes, Y., Pradella, L., & Westwood, S. J. (2024). Correcting misperceptions of partisan opponents is not effective at treating democratic ills. *PNAS Nexus*, 3(8), pgae304. <https://doi.org/10.1093/pnasnexus/pgae304>

Dorison, C. A., & Minson, J. A. (2022). You can't handle the truth! Conflict counterparts overestimate each other's feelings of self-threat. *Organizational Behavior and Human Decision Processes*, 170, 104147.

Dovidio, J. F., Piliavin, J. A., Schroeder, D. A., & Penner, L. A. (2017). *The social psychology of prosocial behavior* (1st ed.). Psychology Press.  
<https://doi.org/https://doi.org/10.4324/9781315085241>

Druckman, J. N., & Levendusky, M. S. (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly*, 83(1), 114–122.

Duarte, J. L., Crawford, J. T., Stern, C., Haidt, J., Jussim, L., & Tetlock, P. E. (2015). Political diversity will improve social psychological science1. *Behavioral and Brain Sciences*, 38, e130.

Epley, N., Kardas, M., Zhao, X., Atir, S., & Schroeder, J. (2022). Undersociality: miscalibrated social cognition can inhibit social connection. *Trends in Cognitive Sciences*, 26(5), 406–418.

Epley, N., Keysar, B., Boven, L. Van, & Gilovich, T. (2004). Perspective taking as egocentric

- anchoring and adjustment. *Journal of Personality and Social Psychology*, 87(3), 327–339.
- Epley, N., & Schroeder, J. (2014). Mistakenly seeking solitude. *Journal of Experimental Psychology: General*, 143(5), 1980.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160.
- Fernbach, P. M., & Van Boven, L. (2022). False polarization: Cognitive mechanisms and potential solutions. *Current Opinion in Psychology*, 43, 1–6.  
<https://doi.org/https://doi.org/10.1016/j.copsyc.2021.06.005>
- Finkel, E. J., Slotter, E. B., Luchies, L. B., Walton, G. M., & Gross, J. J. (2013). A brief intervention to promote conflict reappraisal preserves marital quality over time. *Psychological Science*, 24(8), 1595–1601.
- Fujita, K., Trope, Y., Liberman, N., & Levin-Sagi, M. (2006). Construal levels and self-control. *Journal of Personality and Social Psychology*, 90(3), 351.
- Gaertner, S. L., Dovidio, J. F., Guerra, R., Hehman, E., & Saguy, T. (2015). A common ingroup identity: Categorization, identity, and intergroup relations. In *Handbook of Prejudice, Stereotyping, and Discrimination* (pp. 433–454). Psychology Press.
- Gilbert, D. T. (2002). Inferential correction. In *Heuristics and biases: The psychology of intuitive judgment* (pp. 167–184). Cambridge University Press.
- Gilovich, T., Kruger, J., & Medvec, V. H. (2002). The spotlight effect revisited: Overestimating the manifest variability of our actions and appearance. *Journal of Experimental Social*

*Psychology*, 38(1), 93–99.

- Gilovich, T., Kruger, J., & Savitsky, K. (1999). Everyday egocentrism and everyday interpersonal problems. In *The social psychology of emotional and behavioral problems: Interfaces of social and clinical psychology* (pp. 69–95). American Psychological Association.
- Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgment: an egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology*, 78(2), 211.
- Glynn, C. J., Hayes, A. F., & Shanahan, J. (1997). Perceived support for one's opinions and willingness to speak out: A meta-analysis of survey studies on the "spiral of silence". *Public Opinion Quarterly*, 452–463.
- Goldenberg, A., Abruzzo, J. M., Huang, Z., Schöne, J., Bailey, D., Willer, R., Halperin, E., & Gross, J. J. (2023). Homophily and acrophily as drivers of political segregation. *Nature Human Behaviour*, 7(2), 219–230.
- Golubickis, M., Tan, L. B. G., Falben, J. K., & Macrae, C. N. (2016). The observing self: Diminishing egocentrism through brief mindfulness meditation. *European Journal of Social Psychology*, 46(4), 521–527. [https://doi.org/https://doi.org/10.1002/ejsp.2186](https://doi.org/10.1002/ejsp.2186)
- Greenwald, A. G., & Albert, R. D. (1968). Acceptance and recall of improvised arguments. *Journal of Personality and Social Psychology*, 8.
- Grossmann, I., Dorfman, A., Oakes, H., Santos, H. C., Vohs, K. D., & Scholer, A. A. (2021). Training for Wisdom: The Distanced-Self-Reflection Diary Method. *Psychological Science*,

32(3), 381–394. <https://doi.org/10.1177/0956797620969170>

- Grossmann, I., Gerlach, T. M., & Denissen, J. J. A. (2016). Wise reasoning in the face of everyday life challenges. *Social Psychological and Personality Science*, 7(7), 611–622.
- Grossmann, I., & Kross, E. (2014). Exploring Solomon’s paradox: Self-distancing eliminates the self-other asymmetry in wise reasoning about close relationships in younger and older adults. *Psychological Science*, 25(8), 1571–1580.
- Gunther Moor, B., Crone, E. A., & van der Molen, M. W. (2010). The heartbrake of social rejection: Heart rate deceleration in response to unexpected peer rejection. *Psychological Science*, 21(9), 1326–1333.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon/Random House.
- Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78(1), 81–91. <https://doi.org/10.1037/0022-3514.78.1.81>
- Hawkins, S., Yudkin, D., Juan-Torres, M., & Dixon, T. (2018). *Hidden Tribes: A Study of America’s Polarized Landscape*. New York: More in Common.
- Heltzel, G., & Laurin, K. (2021). Seek and ye shall be fine: Attitudes toward political-perspective seekers. *Psychological Science*, 32(11), 1782–1800.
- Henrich, J. (2015). Culture and social behavior. *Current Opinion in Behavioral Sciences*, 3, 84–89.
- Hogg, M. A. (2016). *Social identity theory*. Springer.

- Hussein, M. A., & Wheeler, S. C. (2024). Reputational costs of receptiveness: When and why being receptive to opposing political views backfires. *Journal of Experimental Psychology: General*.
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3), 405–431.
- Jetten, J., & Hornsey, M. J. (2014). Deviance and dissent in groups. *Annual Review of Psychology*, 65(1), 461–485.
- Kardas, M., Kumar, A., & Epley, N. (2024). Let it go: How exaggerating the reputational costs of revealing negative information encourages secrecy in relationships. In *Journal of Personality and Social Psychology* (Vol. 126, Issue 6, pp. 1052–1083). American Psychological Association. <https://doi.org/10.1037/pspi0000441>
- Kossowska, M., Czarnek, G., & Szwed, P. (2023). Political ideology and belief change in the face of counterevidence. *European Journal of Social Psychology*, 53(6), 1157–1171.
- Kross, E., & Ayduk, O. (2017). Self-distancing: Theory, research, and current directions. In *Advances in Experimental Social Psychology* (Vol. 55, pp. 81–136). Elsevier.
- Kross, E., & Ayduk, O. (2024). Distancing: What it is, how it works, and where to go next. In *Handbook of Emotion Regulation* (3rd ed., pp. 505–512). The Guilford Press.
- Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., Bremner, R., Moser, J., & Ayduk, O. (2014). Self-talk as a regulatory mechanism: how you do it matters. *Journal of Personality and Social Psychology*, 106(2), 304.
- Kross, E., & Grossmann, I. (2012). Boosting wisdom: distance from the self enhances wise

reasoning, attitudes, and behavior. *Journal of Experimental Psychology: General*, 141(1), 43–48.

Kruglanski, A. W., Pierro, A., Mannetti, L., & Grada, E. De. (2006). Groups as Epistemic Providers: Need for Closure and the Unfolding of Group-Centrism. *Psychological Review*, 113(1), 84–100.

Kruglanski, A. W., & Webster, D. M. (1991). Group members' reactions to opinion deviates and conformists at varying degrees of proximity to decision deadline and of environmental noise. *Journal of Personality and Social Psychology*, 61(2), 212–225.

Leary, M. R., & Downs, D. L. (1995). Interpersonal functions of the self-esteem motive: The self-esteem system as a sociometer. In *Efficacy, agency, and self-esteem* (pp. 123–144). Springer.

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, 4(3), 279–286.

Levendusky, M. S., & Malhotra, N. (2016). (Mis) perceptions of partisan polarization in the American public. *Public Opinion Quarterly*, 80(S1), 378–391.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, 126, 88–106.

Libby, L. K., & Eibach, R. P. (2011). Visual perspective in mental imagery: A representational tool that functions in judgment, emotion, and self-insight. In *Advances in experimental social psychology* (Vol. 44, pp. 185–245). Elsevier.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings.

*Psychological Bulletin*, 127(2), 267.

MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin*, 131(2), 202–223.

<https://doi.org/10.1037/0033-2909.131.2.202>

Mason, L., & Wronski, J. (2018). One tribe to bind them all: How our social group attachments strengthen partisanship. *Political Psychology*, 39, 257–277.

Matthes, J., Knoll, J., & von Sikorski, C. (2017). The “spiral of silence” revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression. *Communication Research*, 45(1), 3–33.

<https://doi.org/10.1177/0093650217745429>

Mernyk, J. S., Pink, S. L., Druckman, J. N., & Willer, R. (2022). Correcting inaccurate metaperceptions reduces Americans’ support for partisan violence. *Proceedings of the National Academy of Sciences*, 119(16), e2116851119.

Moore, M., Dorison, C. A., & Minson, J. A. (2023). The contingent reputational benefits of selective exposure to partisan information. In *Journal of Experimental Psychology: General* (Vol. 152, Issue 12, pp. 3490–3525). American Psychological Association.

<https://doi.org/10.1037/xge0001463>

Motyl, M., Iyer, R., Oishi, S., Trawalter, S., & Nosek, B. A. (2014). How ideological migration geographically segregates groups. *Journal of Experimental Social Psychology*, 51, 1–14.

Moy, P., Domke, D., & Stamm, K. (2001). The spiral of silence and public opinion on affirmative action. *Journalism & Mass Communication Quarterly*, 78(1), 7–25.



- Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University.
- Nemeth, Charlan J., & Goncalo, J. A. (2011). Rogues and heroes: Finding value in dissent. *Rebels in Groups: Dissent, Deviance, Difference and Defiance*. Chichester, UK: Blackwell Publishing, 17–35.
- Nemeth, Charlan Jeanne. (1995). Dissent as driving cognition, attitudes, and judgments. *Social Cognition*, 13(3), 273–291.
- Nemeth, Charlan Jeanne, Connell, J. B., Rogers, J. D., & Brown, K. S. (2001). Improving decision making by means of dissent 1. *Journal of Applied Social Psychology*, 31(1), 48–58.
- Nickerson, R. S. (1999). How we know and sometimes misjudge what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, 125(6), 737–759.
- Noelle-Neumann, E. (1993). *The spiral of silence: Public opinion--Our social skin*. University of Chicago Press.
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330.
- OpenAI. (2024). *ChatGPT* (July 2024). <https://chat.openai.com/chat>
- Piaget, J. (1926). *The child's conception of the world*. Routledge and Keegan Paul.
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28(3), 369–381.
- Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and

- misunderstanding. In *Values and knowledge* (1st ed., pp. 103–135). Psychology Press.
- Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., Barea-Arroyo, P., Berge, M. L., Bjørndal, L. D., Bursalıoğlu, A., Bühler, V., Čadek, M., Çetinçelik, M., Clay, G., Cortijos-Bernabeu, A., Damnjanović, K., Dugue, T. M., Esberg, M., Esteban-Serna, C., ... Folke, T. (2021). The general fault in our fault lines. *Nature Human Behaviour*, 5(10), 1369–1380. <https://doi.org/10.1038/s41562-021-01092-x>
- Schachter, S. (1951). Deviation, rejection, and communication. *The Journal of Abnormal and Social Psychology*, 46(2), 190.
- Scheufele, D. A., Shanahan, J., & Lee, E. (2001). Real talk: Manipulating the dependent variable in spiral of silence research. *Communication Research*, 28(3), 304–324.
- Staudinger, U. M., & Glück, J. (2011). Psychological wisdom research: Commonalities and differences in a growing field. *Annual Review of Psychology*, 62, 215–241.
- Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, 62(5), 787.
- Swann, W. B., Jetten, J., Gómez, Á., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review*, 119(3), 441–456.
- Tappin, B. M., Berinsky, A. J., & Rand, D. G. (2023). Partisans' receptivity to persuasive messaging is undiminished by countervailing party leader cues. *Nature Human Behaviour*, 7(4), 568–582.
- Thomas, G., & Fletcher, G. J. O. (2003). Mind-reading accuracy in intimate relationships:

- assessing the roles of the relationship, the target, and the judge. *Journal of Personality and Social Psychology*, 85(6), 1079.
- Todd, A. R., Forstmann, M., Burgmer, P., Brooks, A. W., & Galinsky, A. D. (2015). Anxious and egocentric: How specific emotions influence perspective taking. *Journal of Experimental Psychology. General*, 144(2), 374–391.
- Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, 110(3), 403–421.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124–1131.
- Van Bavel, J., & Packer, D. J. (2021). *The Power of Us: Harnessing Our Shared Identities for Personal and Collective Success*. Hachette UK.
- Van Bavel, J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, 22(3), 213–224.
- Wald, K. A., Kardas, M., & Epley, N. (2024). Misplaced divides? Discussing political disagreement with strangers can be unexpectedly positive. *Psychological Science*, 35(5), 471–488.
- White, R. E., Kuehn, M. M., Duckworth, A. L., Kross, E., & Ayduk, Ö. (2019). Focusing on the future from afar: Self-distancing from future stressors facilitates adaptive coping. *Emotion*, 19(5), 903–916.
- Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, 58, 425–452.
- Willroth, E. C., & Atherton, O. E. (2024). Best laid plans: A guide to reporting preregistration deviations. *Advances in Methods and Practices in Psychological Science*, 7(1),

25152459231213800.

Xu, M., & Petty, R. E. (2022). Two-sided messages promote openness for morally based attitudes. *Personality and Social Psychology Bulletin*, 48(8), 1151–1166.

Yudkin, D., Hawkins, S., & Dixon, T. (2019). *The perception gap: How false impressions are pulling Americans apart*. More In Common.

Zimmerman, F., Garbulsky, G., Ariely, D., Sigman, M., & Navajas, J. (2022). Political coherence and certainty as drivers of interpersonal liking over and above similarity. *Science Advances*, 8(6), eabk1909.

Zuckerman, M., Kernis, M. H., Guarnera, S. M., Murphy, J. F., & Rappoport, L. (1983). The egocentric bias: Seeing oneself as cause and target of others' behavior. *Journal of Personality*, 51(4), 621–630.