

Whoever Is Not With Me Is Against Me: The “Moderate as Out-Group” Effect

Giulia Maimone¹ and Craig R. M. McKenzie^{2, 3}

¹Warrington College of Business, University of Florida

²Rady School of Management, University of California, San Diego

³Department of Psychology, University of California, San Diego

Common intuition suggests that expressing moderate views would allow people to appeal to the broadest audience possible. But is that really the case? Do moderates please all sides or please no side? Across five preregistered studies ($N = 3,272$), we show that people holding a partisan view on a sociopolitical issue perceive moderates (i.e., people disclosing a genuinely nonextreme position) as belonging to the out-group ideology. We find that the “moderate as out-group” effect occurs when sociopolitical issues are moralized and, at the same time, the opposing side is perceived to be a threat to oneself, close ones, or society at large. We also present evidence that the effect is due to partisans’ perception that moderates lack out-group hate, rather than lack in-group love. In other words, partisans perceive that moderates agree with—or simply do not condemn—the opposing (immoral and threatening) ideology, rather than disagree with their own. Consistent with this, the “moderate as out-group” effect occurs when the moderate view is framed as pro-both sides, but it is attenuated when it is framed as pro-neither side.

Public Significance Statement

A wealth of research has investigated how people on opposite sides of the heavily polarized political spectrum judge one another. However, how partisans judge and categorize moderates remains unanswered. Understanding when and why expressing moderate views pleases all sides versus no side is essential to navigating the complex and increasingly hostile sociopolitical world. We present a novel experimental paradigm for studying interpersonal categorization, showing that partisans tend to categorize moderates as out-group members, especially when sociopolitical issues are moralized and the opposing side is perceived as threatening. Finally, we show that the effect is driven by the perception that moderates fail to repudiate the (immoral and threatening) out-group. This work speaks to research in decision making, social cognition, and political theory.

Keywords: moderate, polarization, out-group, moral, threat

Supplemental materials: <https://doi.org/10.1037/xge0001957.supp>

In recent years, political polarization in the United States has reached unprecedented levels (Finkel et al., 2020). A wealth of research has examined how people on opposite sides of the political spectrum perceive one another (e.g., Hartman et al., 2023; Iyengar et al., 2019; Schwalbe et al., 2020; Tappin & McKay, 2019), as well as strategies to

reduce the psychological distance between opposing partisans (e.g., Finkel et al., 2020; Lees & Cikara, 2020; Minson & Chen, 2022; Pennycook et al., 2020). Substantially less attention has been devoted to understanding how partisans perceive and judge those who do not align with either ideological camp.

Eddie Brummelman served as action editor.

Giulia Maimone  <https://orcid.org/0000-0003-3927-7157>

All preregistrations, data, code, and materials are available on ResearchBox at <https://researchbox.org/3048>. This article was presented at the 3rd annual California School Conference; at the 55th annual conference of the Association for Consumer Research; at the 2024 annual conference of the Norms and Behavioral Change Center at the University of Pennsylvania; at the 30th biannual conference on Subjective Probability, Utility, and Decision Making; at the 46th annual conference of the Society for Judgment and Decision Making; and at the 2025 MIT Directions of Polarization, Social Norms and Trust in Societies Workshop.

The authors do not have organizational affiliations, stock ownership, patent filings, public statements and positions, or additional information to disclose. The authors have research support to disclose: National Science

Foundation (Grant SES-2049935) to Craig R. M. McKenzie. The authors thank Mohamed Hussein and Joachim Vosgerau for having commented on an early version of the article and Eugene Caruso, Sahar Fard, Uma Kar-markar, and Wendy Liu for helpful comments.

Giulia Maimone played a lead role in data curation, investigation, project administration, visualization, writing—original draft, and writing—review and editing and an equal role in conceptualization and methodology. Craig R. M. McKenzie played a lead role in funding acquisition and resources, a supporting role in investigation and writing—review and editing, and an equal role in conceptualization and methodology.

Correspondence concerning this article should be addressed to Giulia Maimone, Warrington College of Business, University of Florida, 1384 Union Road Bryan, Hall 100, Gainesville, FL 32611, United States. Email: giulia.maimone@warrington.ufl.edu

In this article, we investigate how partisans¹ categorize moderates—that is, individuals who disclose genuinely intermediate, ambivalent, or nonextreme positions (Fowler et al., 2023).² Under this definition, moderates are nonpartisan, as they do express a position (unlike neutral individuals; Silver & Shaw, 2022; Zlatev, 2019), yet do not explicitly endorse either side (unlike ambivalent partisans; Siev et al., 2024). Because moderates occupy an intermediate ideological space, their categorization represents a nontrivial judgment problem. Partisan perceivers may recognize them as moderates, assimilate them into the in-group, or categorize them as out-group members.

Intuition suggests that moderate positions should broaden appeal. Political science research has documented electoral advantages for moderate candidates in general elections (Hall, 2015), and extensive work in psychology has found reputational benefits for those who are receptive to opposing views (e.g., Chen et al., 2010; Heltzel & Laurin, 2021; Hussein & Tormala, 2021). These findings suggest that moderates may be recognized as moderates or even assimilated into the in-group. Contrary to this expectation, we find that partisans systematically categorize moderates as members of the opposing ideological camp. Across five preregistered studies, individuals holding partisan views on a sociopolitical issue—regardless of which side they are on—tend to classify moderate positions, and the individuals expressing them, as belonging to the out-group. For example, in the context of abortion, pro-life individuals categorize moderate positions on abortion as pro-choice, whereas pro-choice individuals categorize the same targets as pro-life. We term this phenomenon the “moderate as out-group” effect.

Extensive research has demonstrated that social perceptions rely on categorization processes that are essential to navigate the social world (e.g., Allport, 1954; Fiske, 1998; Hamilton, 1981; Macrae & Bodenhausen, 2000; Markman & Gentner, 2001; Tajfel, 1969). Categorization processes are shaped by a combination of assimilation and contrast effects (Sherif & Hovland, 1961) and by implicit thresholds that determine whether a target falls within or outside a category. In highly polarized contexts—where extreme out-group members are salient (Finkel et al., 2020)—moderates might be expected to benefit from contrast and thus be evaluated relatively positively (Wilder & Thompson, 1988). However, contemporary political polarization is marked not only by ideological divergence but also by intense intergroup hostility, with Democrats and Republicans deeming each other biased (e.g., Hartman et al., 2023), detestable and untrustworthy (e.g., Iyengar et al., 2019), and even immoral (e.g., Tappin & McKay, 2019). The mutual feelings of disliking and animosity between opposing partisans are known as affective polarization and are described as the gap between positive feelings for the in-group (“in-group love”) and negative feelings for the out-group (“out-group hate”; Iyengar et al., 2012). This suggests that category boundaries may be drawn in ways that exclude intermediate positions rather than assimilate them.

A psychological framework widely invoked to explain intergroup misperception is naïve realism—the belief that one’s own views are objective and that disagreement reflects bias in others (e.g., Griffin & Ross, 1991; Pronin et al., 2002, 2004; Robinson et al., 1995; Ross, 2013; Ross et al., 2010; Ross & Ward, 1995; Schwalbe et al., 2020). Naïve realism has been used to account for phenomena such as the hostile media effect, in which opposing groups perceive identical news coverage as biased against their side (e.g., Feldman, 2014; Giner-Sorolla & Chaiken, 1994; Vallone et al., 1985), as well as false polarization—the tendency to exaggerate the extremity of

opposing partisans (e.g., Fernbach & Van Boven, 2022; Fiorina et al., 2008; Keltner & Robinson, 1996; Lees & Cikara, 2021). At first glance, our finding that moderates are judged as out-group members may appear to be in tension with false polarization research. False polarization typically examines label-to-position inferences—decision-makers begin with a partisan label (e.g., “conservative”) and infer the positions that characterize that group, thereby speaking to prototype representation. In contrast, our paradigm examines position-to-label categorization—decision-makers begin with a specific position and infer the partisan label it elicits, thereby speaking to category boundary placement. Although closely related, these inferential questions are analytically distinct. Recognizing this distinction helps clarify how exaggerated perceptions of out-group prototypes and the categorization of moderates as out-group members can arise within the same broader process of partisan perception—both reflecting a tendency to interpret political information in ways that accentuate intergroup divisions. In this sense, our findings complement—rather than contradict—false polarization by highlighting how partisan boundaries are drawn when targets deviate from partisan orthodoxy. Although naïve realism may plausibly contribute to the “moderate as out-group” effect, our data show that it is unlikely to uniquely account for the phenomenon.

Among the previously discussed tendencies—namely, viewing opposing partisans as biased (e.g., Hartman et al., 2023), detestable (e.g., Iyengar et al., 2019), and immoral (e.g., Tappin & McKay, 2019)—we propose that moralization plays a central role in shaping partisan category boundaries. Moralized issues activate sacred values central to group identity (Durkheim, 1925/1973; Graham et al., 2011; Graham & Haidt, 2012; Haidt, 2007). Because sacred values are experienced as absolute and inviolable (Tetlock, 2003), moralized issues are especially polarizing (Clifford, 2019; Graham & Haidt, 2012), casting the in-group as morally righteous and the out-group as morally deviant (Haidt, 2012), with little room for shades of gray in between. Under such conditions, partisans may not classify moderate positions as moderate, but instead assimilate them to one of the two poles (i.e., the in-group or the out-group). Sacredness is closely tied to purity, a fundamental moral intuition (Haidt, 2001; Haidt & Joseph, 2007), and deviations from sacred values can trigger indignation and reputational costs. Indeed, receptiveness to immoral others elicits condemnation (Hussein & Wheeler, 2024), and partisans expressing ambivalent positions on highly polarizing issues are liked less by both in-group and out-group members (Siev et al., 2024). The moral stain resulting from failing to condemn the “evil” side can be so aversive that partisans prefer harming their in-group over helping the out-group (Gershon & Fridman, 2022), illustrating how the “out-group hate” motive outweighs the “in-group love” motive in intergroup conflicts (e.g., Halevy et al., 2008; Rathje et al., 2021). At the same time, recent philosophical work has suggested that not all groups

¹ The Merriam-Webster dictionary defines the noun “partisan” as “a firm adherent to a party, faction, cause, or person.” In political science, the noun “partisan” is commonly used to refer to Republicans and Democrats (i.e., people who support a party in the U.S. political system). In this research, we use the noun “partisan” to refer to people with partial or unilateral views on a specific sociopolitical issue. For example, in this work, pro-life and pro-choice individuals will be called “partisans” with respect to abortion.

² Fowler et al. (2023) conceptualized three classes of “moderates”: (a) having genuinely moderate views across issues, (b) not caring about politics, and (c) not identifying along the traditional liberal–conservative dimension. In this work, we focus on individuals’ views on a single sociopolitical issue (rather than on their overall political identity), so we adopt only the first class to define “moderates.”

are equally willing to exert effort in support of a moral cause or in opposition to an immoral one. Groups that are—or subjectively feel—threatened by the absence or potential loss of specific rights are more motivated and involved (Fard, 2025). Perceived threat may therefore raise the threshold for in-group inclusion, biasing perceptions of similarity and rendering moderate positions insufficiently aligned with the in-group.

Building on these literatures, we hypothesize that people holding partisan views on moralized (vs. preference-based) issues will be more likely to categorize moderates as out-group members and that partisans who perceive the opposing ideology as threatening (vs. nonthreatening) will also be more likely to do so. We further hypothesize that the “moderate as out-group” effect is driven more by aversion to the moderate target’s perceived agreement with the immoral out-group (reflecting the absence of out-group hate) than by aversion to the moderate’s perceived disagreement with the moral in-group (reflecting the absence of in-group love). Across five preregistered studies, we find support for these hypotheses across a variety of sociopolitical issues and manipulations of moderation.

Empirical Approach

When testing the “moderate as out-group” hypothesis, moderation can be conceptualized in two ways: as the nonextremity of a position or policy or as the nonextremity of an individual who endorses a moderate position. We examine both conceptualizations. Study 1 tests whether partisans perceive moderate policies as belonging to the out-group ideology. Study 2 shows that the findings of Study 1 extend to judgment of individuals supporting such moderate policies. Study 3 identifies the necessary conditions under which the “moderate as out-group” effect emerges, and Studies 4 and 5 provide evidence regarding its underlying psychological mechanism.

Across four of the five studies,³ participants were grouped into five nonrandomized conditions based on their own position on the focal sociopolitical issue—specifically, two more extreme partisan positions (opposed to each other), two less extreme partisan positions (opposed to each other), and a nonpartisan position. For example, in the abortion studies, participants could identify as definitely pro-life, mostly pro-life, neither pro-life nor pro-choice, mostly pro-choice, or definitely pro-choice. Importantly, these categories describe participants’ own ideological stance; they do not refer to the positions participants evaluated. All participants—regardless of their own stance—then evaluated both moderate and nonmoderate targets. We expected the “moderate as out-group” effect to be especially pronounced among participants holding more extreme partisan views. However, to test a conservative and broadly generalizable hypothesis, our primary prediction was that the effect would be observable among partisans overall, regardless of their level of extremity. Accordingly, in our primary analyses, we collapsed across more- and less-extreme partisans within each ideological camp. This yields three between-participants groups: (a) partisans on one side of the issue, (b) partisans on the other side, and (c) nonpartisans. After establishing the effect at this aggregated level, we then conduct secondary analyses using the full five-category classification of participant’s ideology. These analyses allowed us to examine whether—and how—the effect varies as a function of participant extremity.

Our studies differed in how moderate positions were operationalized. In Studies 1 and 2, target positions were manipulated along a linear policy continuum. In Studies 3 and 4, moderation was operationalized

quantitatively using a numerical scale. In Study 5, moderation was manipulated discretely using verbal labels.

Transparency and Openness

All studies were preregistered with directional hypotheses. The preregistrations, de-identified data, analysis code, and experimental materials for all studies are publicly available on ResearchBox (<https://researchbox.org/3048>). Across studies, we reported all manipulations, measures, and data exclusions. All sample sizes and exclusion criteria were determined in advance and specified in the preregistrations. In addition to preregistered analyses, we report additional analyses that are clearly labeled as exploratory. Finally, all statistical tests reported in this article are two-tailed. See the [Supplemental Materials](#) for analyses without exclusions.

Study 1: The “Moderate as Out-Group” Effect—Judgment of Different Policies on Abortion

Study 1 employs a novel experimental paradigm and provides evidence for the “moderate as out-group” effect. Specifically, it finds that individuals holding partisan views on abortion categorized moderate policies on abortion as belonging to the out-group ideology.

Method

Participants

Seven hundred ninety-eight U.S.-based CloudResearch-approved Amazon Mechanical Turk workers completed the experiment in exchange for monetary compensation (\$0.30 for an estimated 2-min completion time). As preregistered, participants who failed an attention-check question were not allowed to start the study. Additionally, participants who provided inconsistent ratings of the policies (e.g., reporting the policy “Abortion never allowed” was pro-choice, “Abortion always allowed” was pro-life, or rating “Abortion allowed up to 4 weeks” as more pro-life than “Abortion never allowed,” “Abortion allowed up to 8 weeks” as more pro-life than “Abortion allowed up to 4 weeks,” and so on for all policies) were excluded from analyses. Our final sample consisted of 579 participants (251 identified as female, 324 identified as male, and four identified as nonbinary; $M_{\text{age}} = 40.4$ years, $SD_{\text{age}} = 11.8$ years).

Procedure

Participants first reported their position on abortion on a 5-point bipolar scale ($-2 = \text{definitely pro-life}$, $-1 = \text{mostly pro-life}$, $0 = \text{neither pro-life nor pro-choice}$, $1 = \text{mostly pro-choice}$, $2 = \text{definitely pro-choice}$). They then evaluated 11 policies on abortion for non-medical reasons ranging from “Abortion never allowed” to “Abortion always allowed” in intervals of 4 weeks (“Abortion allowed up to 4 weeks after conception,” “Abortion allowed up to 8 weeks after conception,” ... “Abortion allowed up to 36 weeks after conception”) and judged each policy on a 7-point bipolar scale ($-3 = \text{policy is definitely pro-life}$, $0 = \text{policy is neither pro-life nor pro-choice}$, $3 = \text{policy is definitely pro-choice}$). Finally, participants reported their political affiliation on a 5-point bipolar scale ($-2 = \text{extremely$

³ Study 4 employs a different paradigm, described below.

conservative, 0 = neither conservative nor liberal, 2 = extremely liberal) and their gender (female, male, other) and age (free-response box). For half of the participants, the order of the policies and the direction of all scales were reversed.

Results and Discussion

We first had to answer the question of what constitutes a moderate policy on abortion. Whereas 98% of countries worldwide allow abortion to save a woman's life, only 34% of countries (accounting for roughly 60% of the world's population) allow abortion upon a woman's request.⁴ The majority of these countries permit abortion on request up to 12 weeks from conception (i.e., within the first trimester), suggesting that abortion allowed within the end of the first trimester could be considered a moderate policy on abortion for nonmedical reasons. Consistent with this, the three policies in our study that allowed abortion up to 4, 8, and 12 weeks from conception were identified by participants as moderate. In fact, participants who reported being neither pro-life nor pro-choice (i.e., nonpartisan participants)⁵ judged these policies as neither pro-life nor pro-choice (i.e., they were rated as not significantly different from 0 on the pro-life/pro-choice scale; one-sample t tests against 0: all $ps > .180$; see the green points for these policies in Panel A of Figure 1).

As preregistered, we first analyzed the judgments of the policies on abortion made by three groups of respondents: pro-life participants (i.e., participants who indicated being "mostly pro-life" or "definitely pro-life"), pro-choice participants (i.e., participants who indicated being "mostly pro-choice" or "definitely pro-choice"), and participants who indicated being neither pro-life nor pro-choice. As noted above, nonpartisan participants judged the three moderate policies as moderate. Supporting our "moderate as out-group" hypothesis, pro-life participants judged all three moderate policies as pro-choice (i.e., rated significantly greater than 0; $M_{\text{UpTo4Weeks}} = 0.60$, $SD = 2.17$; $M_{\text{UpTo8Weeks}} = 1.08$, $SD = 1.98$; $M_{\text{UpTo12Weeks}} = 1.66$, $SD = 1.63$; one-sample t tests against 0: all $ts > 3.30$, $ps < .0012$; see the red points for these policies in Panel A of Figure 1), whereas pro-choice participants judged the same three policies as pro-life (i.e., rated significantly lower than 0; $M_{\text{UpTo4Weeks}} = -1.80$, $SD = 1.48$; $M_{\text{UpTo8Weeks}} = -1.38$, $SD = 1.60$; $M_{\text{UpTo12Weeks}} = -0.73$, $SD = 1.78$; one-sample t tests against 0: all $ts < -8.30$, $ps < .001$; see the blue points for these policies in Panel A of Figure 1).

We then considered the more granular position on abortion we collected; that is, we analyzed the judgments of five groups of respondents: "definitely pro-life," "mostly pro-life," "neither pro-life nor pro-choice," "mostly pro-choice," and "definitely pro-choice." We find that the "moderate as out-group" effect is more pronounced for participants who reported more extreme views on abortion. "Definitely pro-life" participants judged the three moderate policies as more pro-choice ($M_{\text{UpTo4Weeks}} = 1.60$, $SD = 1.92$; $M_{\text{UpTo8Weeks}} = 1.84$, $SD = 1.76$; $M_{\text{UpTo12Weeks}} = 2.16$, $SD = 1.48$) than "mostly pro-life" participants did ($M_{\text{UpTo4Weeks}} = -0.49$, $SD = 1.88$; $M_{\text{UpTo8Weeks}} = 0.26$, $SD = 1.88$; $M_{\text{UpTo12Weeks}} = 1.12$, $SD = 1.62$; all $ts > 4.02$, $ps < .001$; see the red and olive-green points for these policies in Panel B of Figure 1). Similarly, "definitely pro-choice" participants judged the same three policies as more pro-life ($M_{\text{UpTo4Weeks}} = -2.00$, $SD = 1.40$; $M_{\text{UpTo8Weeks}} = -1.58$, $SD = 1.57$; $M_{\text{UpTo12Weeks}} = -0.92$, $SD = 1.78$) than "mostly pro-choice" participants did ($M_{\text{UpTo4Weeks}} = -1.23$, $SD = 1.57$; $M_{\text{UpTo8Weeks}} = -0.77$, $SD = 1.55$; $M_{\text{UpTo12Weeks}} = -0.19$, $SD = 1.65$; all $ts < -3.80$,

$ps < .001$; see the pink and turquoise points for these policies in Panel B of Figure 1).

Beyond the preregistered analysis, we observed that pro-life and pro-choice participants judged all policies—not just moderate ones—differently (all $ts > 11.00$, $ps < .001$), except for "Abortion never allowed" and "Abortion always allowed" ($ps > .058$; see the red and blue points in Panel A of Figure 1). To investigate whether the "moderate as out-group" effect is simply part of a broader pattern (e.g., false polarization; Keltner & Robinson, 1996) or a distinct phenomenon, we created a "policy type" variable with two levels: "moderate" (average judgments for the three moderate policies, i.e., abortion allowed up to 4, 8, and 12 weeks) and "not moderate" (average judgments for all other policies). Restricting the dataset to participants who expressed a partisan position ($N = 553$), we then ran a non-preregistered linear model regressing judgment on policy type (moderate vs. not moderate), participants' orientation on abortion (pro-life vs. pro-choice), and their interaction, clustering the standard errors (SEs) by participant. Predictors were effect coded ($-1/+1$). This model revealed a significant main effect of orientation, such that pro-life participants judged all policies as more pro-choice than pro-choice participants did and vice versa ($b = 0.89$, $t = 15.88$, $p < .001$), and a significant main effect of policy type, such that nonmoderate policies were judged as more pro-choice⁶ than moderate ones ($b = 0.71$, $t = 20.35$, $p < .001$). Finally, it revealed a significant interaction showing that opposing partisans judged moderate policies more differently than nonmoderate policies ($b = 0.32$, $t = 9.12$, $p < .001$; see Figure 2). These results show that opposing partisans tend to judge all policies differently, but the divergence is greater for moderate ones. Thus, the "moderate as out-group" effect is not merely part of a broader pattern of intergroup perceptions but rather reflects a distinctive flexibility in categorizing moderate policies.

In sum, Study 1 demonstrates that pro-life individuals judged moderate policies on abortion as pro-choice, and pro-choice individuals judged the same moderate policies as pro-life. Individuals with partisan views tend to categorize moderate policies as aligned with the out-group ideology. In the next study, we investigate whether this effect extends to judgments of moderate individuals.

Study 2: The "Moderate as Out-Group" Effect—Judgment of a Person Supporting Different Policies on Abortion

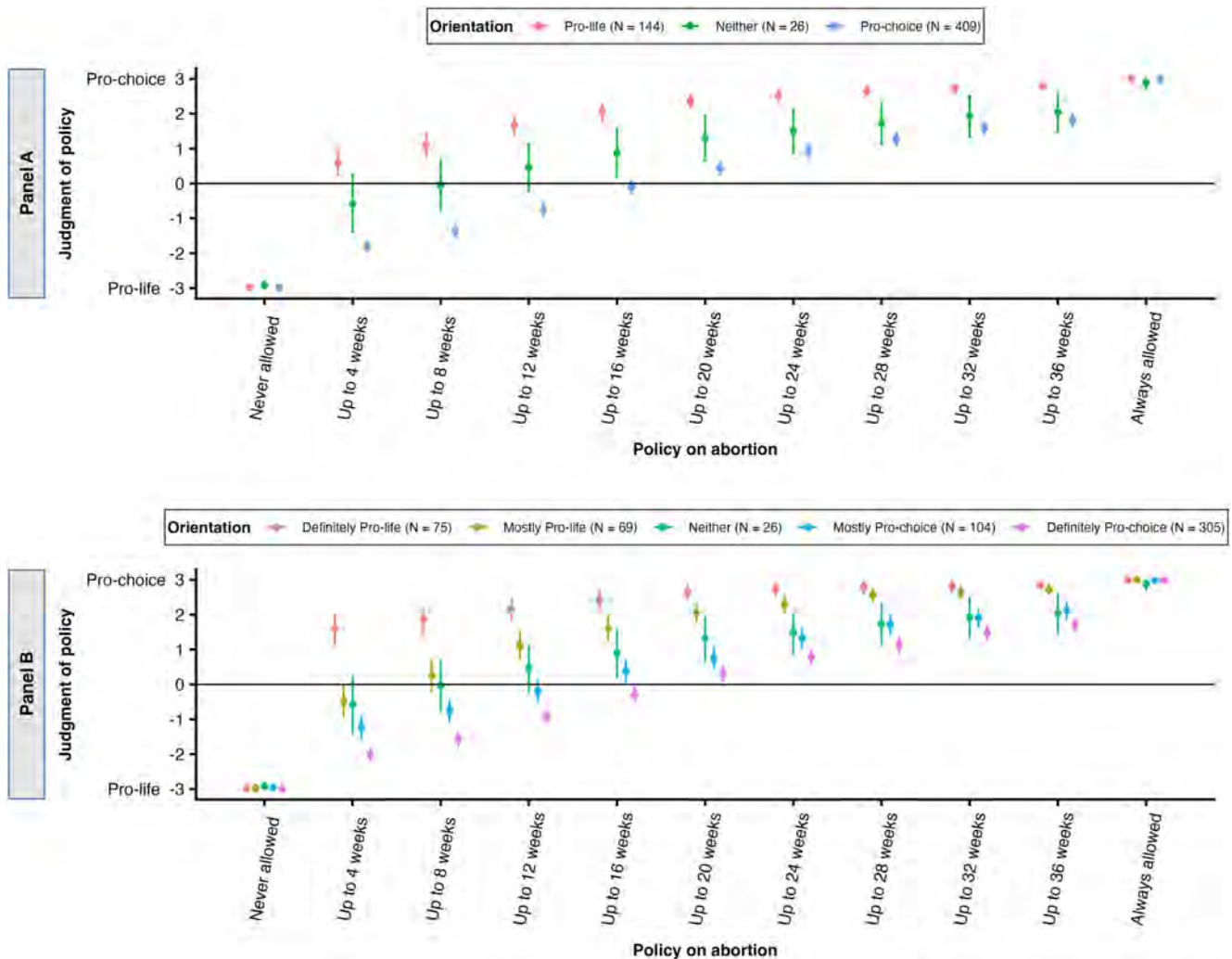
Study 2 employs the same design and experimental procedure as Study 1, except that, rather than judging each policy, participants judged a hypothetical person, Jamie, who supported each policy

⁴ United Nations Department of Economic and Social Affairs, World Population Policies 2017: Abortion Laws and Policies (2020).

⁵ Note that we used judgments from nonpartisan participants to define the moderate policies because they should not exhibit any partisan bias, if there is any. Also, it is not relevant to our theorizing to define the exact stance of these nonpartisan participants (e.g., whether they are moderate, indifferent about abortion, or undecided). We only require that they are nonpartisan to identify moderate policies. Note that these are the *judges*, not the targets to be judged. The aim of the study is to test how partisan participants judge moderate policies.

⁶ This is because the "nonmoderate policies" level includes the following policies: abortion never allowed; abortion allowed up to 16, 20, 24, 28, 32, and 36 weeks; and abortion always allowed. Of these policies, only one is pro-life—all the others are pro-choice.

Figure 1
Study 1: “Moderate as Out-Group” Effect



Note. The figure shows the results of Study 1. Panel A shows the policy judgments by three groups (pro-life vs. neither vs. pro-choice). Panel B shows the policy judgments by five groups (definitely pro-life vs. mostly pro-life vs. neither vs. mostly pro-choice vs. definitely pro-choice). Abortion allowed up to 4, 8, and 12 weeks were identified as moderate policies. Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

(one at a time). This study shows that the “moderate as out-group” effect also extends to judgments of moderate people.

Method

Participants

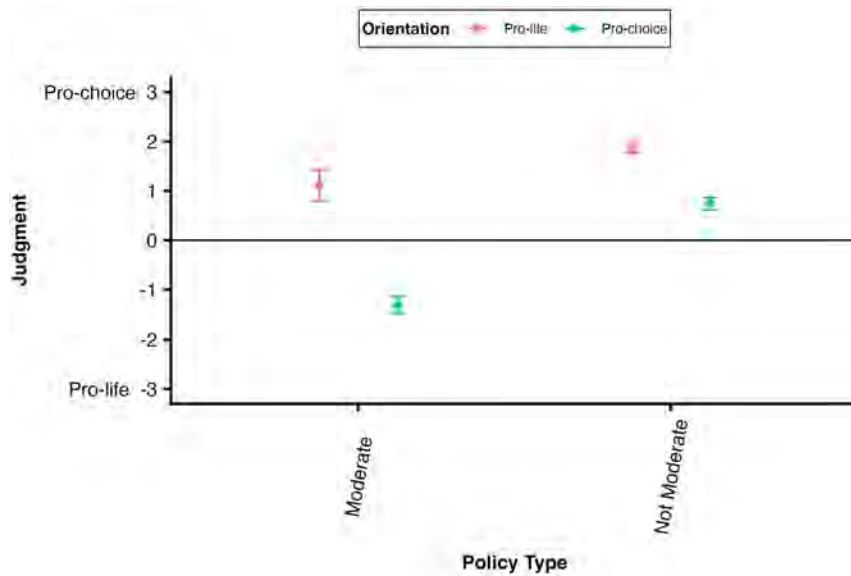
Seven hundred ninety-nine CloudResearch-approved Amazon Mechanical Turk workers from the United States completed the experiment in exchange for monetary compensation (\$0.30 for an expected 2-min completion time). As preregistered, participants who failed to correctly answer an attention-check question were

not allowed to start the study, and those who provided inconsistent ratings of Jamie (using the same parameters as in Study 1) were excluded from the analyses. The final sample consisted of 612 participants (310 identified as female, 294 identified as male, and eight identified as nonbinary; $M_{\text{age}} = 42.0$ years, $SD_{\text{age}} = 12.7$ years).

Procedure

The experimental procedure was the same as in Study 1, except that participants judged a hypothetical person, Jamie. As in Study 1, participants reported their position on abortion and considered the

Figure 2
Study 1: Judgment Gap for Moderate and Nonmoderate Policies



Note. The figure shows the results of Study 1. Policy judgments by participants' orientation on abortion (pro-life vs. pro-choice) and policy type (moderate vs. not moderate). Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

same 11 policies on abortion for nonmedical reasons.⁷ Finally, they judged Jamie on a 7-point bipolar scale ($-3 = \text{Jamie is definitely pro-life}$, $3 = \text{Jamie is definitely pro-choice}$) for supporting (one at a time) each policy. All scales and the order of the policies were reversed for half of the participants.

Results and Discussion

We first analyzed the judgments of Jamie made by three groups of respondents: pro-life (i.e., participants who indicated being “mostly pro-life” or “definitely pro-life”), pro-choice (i.e., participants who indicated being “mostly pro-choice” or “definitely pro-choice”), and participants who indicated being neither pro-life nor pro-choice. Nonpartisan participants judged Jamie as neither pro-life nor pro-choice when Jamie supported policies allowing abortion up to 4, 8, and 12 weeks (one-sample t tests against 0: all $ps > .095$; see the green points for these policies in Panel A of Figure 3). These results replicate the finding from Study 1. Pro-life participants judged Jamie supporting two of the three moderate policies as pro-choice (i.e., rated significantly greater than 0; $M_{\text{UpTo8Weeks}} = 0.61$, $SD = 2.00$; $M_{\text{UpTo12Weeks}} = 1.04$, $SD = 1.85$; one-sample t tests against 0: $ts > 3.95$, $ps < .001$; see the red points for these policies in Panel A of Figure 3).⁸ By contrast, pro-choice participants judged Jamie supporting all three moderate policies as pro-life (i.e., rated significantly lower than 0; $M_{\text{UpTo4Weeks}} = -1.44$, $SD = 1.67$; $M_{\text{UpTo8Weeks}} = -0.95$, $SD = 1.75$; $M_{\text{UpTo12Weeks}} = -0.33$, $SD = 1.84$; one-sample t tests against 0: all $ts < -3.65$, $ps < .001$; see the blue points for these policies in Panel A of Figure 3).

As in Study 1, the “moderate as out-group” effect was more pronounced among participants with more extreme views. “Definitely pro-life” participants judged Jamie supporting all three moderate policies as more pro-choice ($M_{\text{UpTo4Weeks}} = 1.06$, $SD = 2.14$;

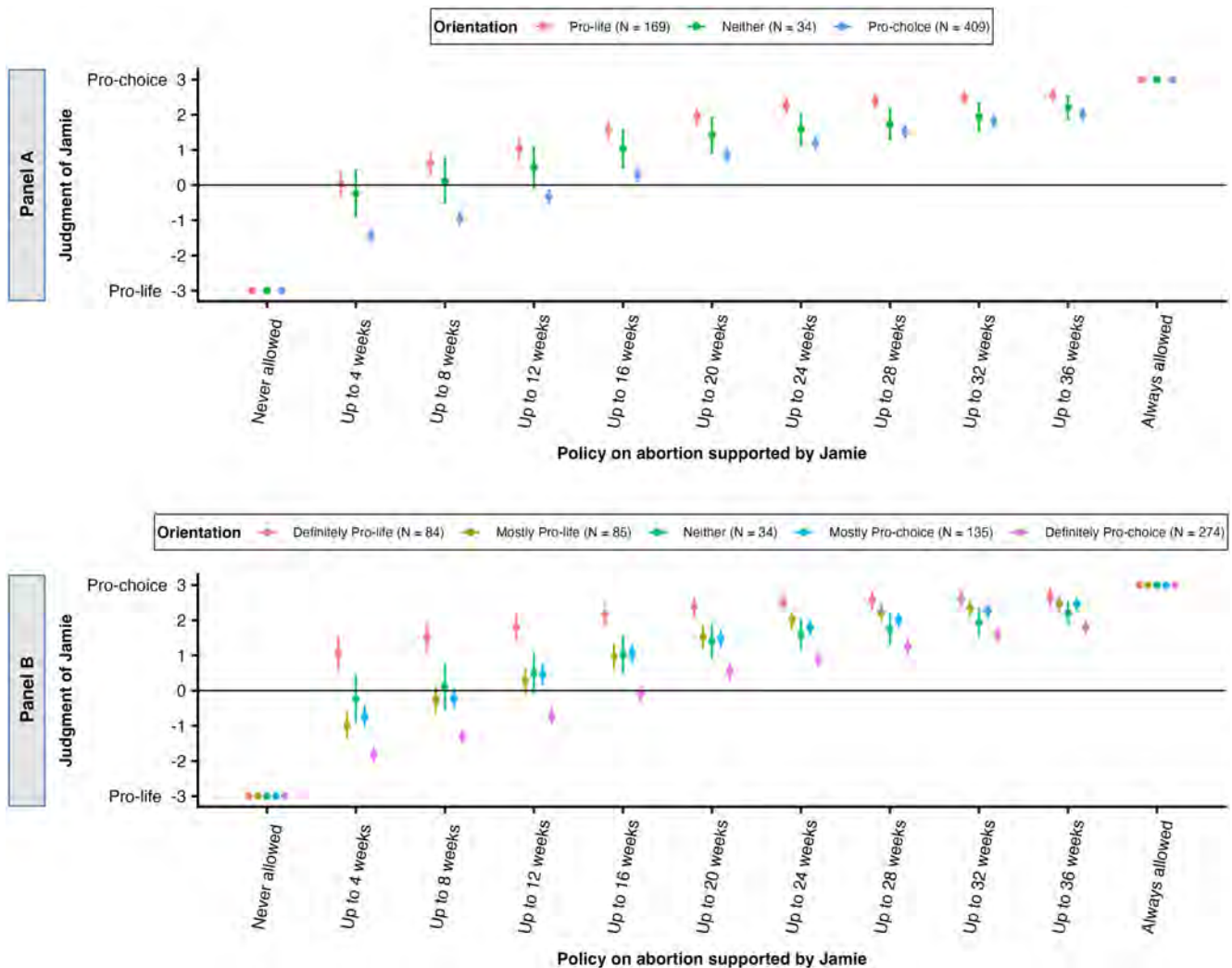
$M_{\text{UpTo8Weeks}} = 1.50$, $SD = 1.91$; $M_{\text{UpTo12Weeks}} = 1.81$, $SD = 1.73$) than “mostly pro-life” participants did ($M_{\text{UpTo4Weeks}} = -0.99$, $SD = 1.67$; $M_{\text{UpTo8Weeks}} = -0.27$, $SD = 1.67$; $M_{\text{UpTo12Weeks}} = 0.27$, $SD = 1.64$; all $ts > 5.92$, $ps < .001$; see the red and olive-green points for these policies in Panel B of Figure 3). Similarly, “definitely pro-choice” participants judged Jamie supporting the same three policies as more pro-life ($M_{\text{UpTo4Weeks}} = -1.80$, $SD = 1.54$; $M_{\text{UpTo8Weeks}} = -1.31$, $SD = 1.71$; $M_{\text{UpTo12Weeks}} = -0.73$, $SD = 1.83$) than “mostly pro-choice” participants did ($M_{\text{UpTo4Weeks}} = -0.73$, $SD = 1.71$; $M_{\text{UpTo8Weeks}} = -0.21$, $SD = 1.60$; $M_{\text{UpTo12Weeks}} = 0.46$, $SD = 1.60$; all $ts < -6.09$, $ps < .001$; see the pink and turquoise points for these policies in Panel B of Figure 3).

As in Study 1, pro-life and pro-choice participants' judgments of Jamie differed for every single policy (all $ts > 5.00$, $ps < .001$), except for “Abortion never allowed” and “Abortion always allowed” ($ps > .317$; see the red and blue points in Panel A of Figure 3). For this reason, we created the same “policy type” variable (moderate vs. not moderate) as in Study 1, averaging judgments of Jamie supporting all moderate versus nonmoderate policies. Restricting the data set to participants who expressed a partisan position ($N = 578$),

⁷ It is important to clarify how abortion policies were operationalized in Studies 1 and 2. Participants evaluated policies spanning the full pregnancy timeline, which were presented as hypothetical anchors along a continuum to examine how partisans categorize positions at varying degrees of permissiveness. Those scenarios were not intended to imply that any real-world actor is currently advocating for policies such as abortion allowed up to 4 or 36 weeks nor to suggest that such positions are common in the public debate. Importantly, the practical feasibility or frequency of these policies is orthogonal to our research question.

⁸ Pro-life participants rated Jamie supporting “Abortion allowed up to 4 weeks from conception” as neither pro-life nor pro-choice: $M_{\text{UpTo4Weeks}} = 0.03$, $SD = 2.17$; one-sample t test against 0: $t(168) = 0.18$, $p = .860$.

Figure 3
Study 2: “Moderate as Out-Group” Effect



Note. The figure shows the results of Study 2. Panel A shows the judgments of Jamie supporting each policy by three groups (pro-life vs. neither vs. pro-choice). Panel B shows the judgments of Jamie by five groups (definitely pro-life vs. mostly pro-life vs. neither vs. mostly pro-choice vs. definitely pro-choice). Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

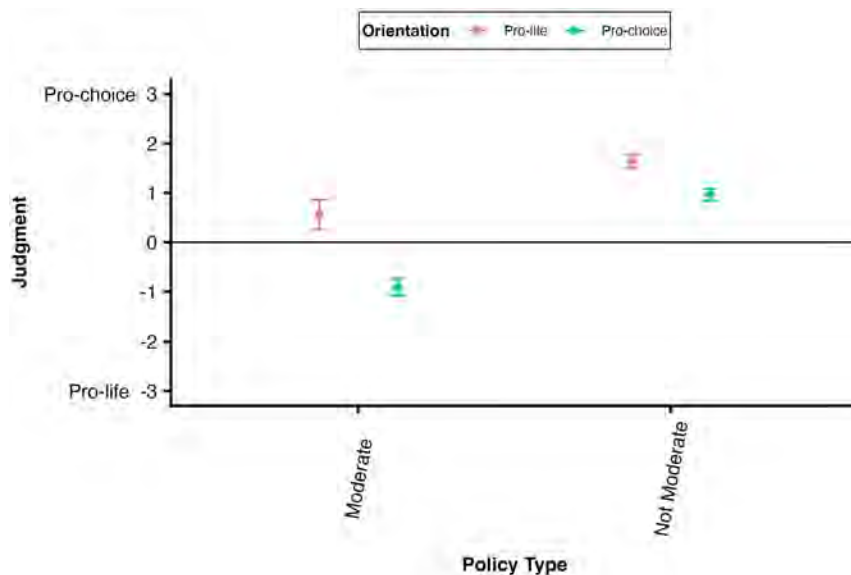
we conducted a non-preregistered linear model regressing judgment of Jamie on policy type (moderate vs. not moderate), participants' orientation on abortion (pro-life vs. pro-choice), and their interaction, clustering the SEs by participant. Predictors were effect coded ($-1/+1$). Mirroring the results of Study 1, this model revealed a significant main effect of orientation, such that pro-life participants judged Jamie supporting all policies as more pro-choice than pro-choice participants did and vice versa ($b = 0.54$, $t = 9.08$, $p < .001$), and a significant main effect of policy type, such that Jamie supporting nonmoderate policies was judged as more pro-choice than when supporting moderate ones ($b = 0.74$, $t = 22.42$, $p < .001$). Finally, it again revealed a significant interaction indicating that opposing partisans exhibited a greater judgment gap when Jamie supported moderate policies on abortion than when they supported nonmoderate ones ($b = 0.19$, $t = 5.90$, $p < .001$; see Figure 4).

Study 2 replicates and extends the findings from Study 1, showing that pro-life participants judged a person endorsing moderate abortion policies as pro-choice, and pro-choice participants judged the same target endorsing the same moderate policies as pro-life. Thus, the “moderate as out-group” effect emerges not only when judging policies or viewpoints but also when judging people. As in Study 1, opposing partisans judged Jamie differently across the entire policy spectrum, but the gap was larger for moderate policies, further underscoring a distinctive flexibility in how moderates are categorized.

Study 3: Necessary Conditions for the “Moderate as Out-Group” Effect

Overall, the results of Studies 1 and 2 demonstrate that individuals with partisan views judge moderate policies and people as belonging

Figure 4
Study 2: Judgment Gap for Moderate and Nonmoderate Person



Note. The figure shows the results of Study 2. Judgment of Jamie supporting policies on abortion by participants' orientation on abortion (pro-life vs. pro-choice) and policy type (moderate vs. not moderate). Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

to the out-group ideology, regardless of which side they are on, supporting the “moderate as out-group” hypothesis. These studies, however, do not identify the conditions under which the effect emerges. Study 3 addresses this gap by testing our hypothesized necessary conditions: moralization of the issue and perceived threat from the opposing side. To do so, we shift from abortion to support for presidential candidates in the 2024 elections. Study 3 was run on October 31, 2024—5 days before the 2024 U.S. presidential election—and concerns support for the two leading candidates, Donald Trump and Kamala Harris. In this study, we assume that most liberals would support Kamala Harris and most conservatives would support Donald Trump. This assumption is grounded in extensive political science research showing that ideological orientation is a key determinant of both partisan alignment (Democrats vs. Republicans) and its intensity (Abramowitz & Saunders, 2006; Bafumi & Shapiro, 2009; Levendusky, 2009). Moreover, recent work suggests that presidential vote choice is increasingly driven by partisan identity, a trend fueled by accelerated social sorting (Vidal & Hunt, 2026).

In Study 3, we also operationalized moderation differently from Studies 1 and 2—quantitatively, as the midpoint of a 7-point bipolar scale (i.e., 0) rather than along a policy continuum. By measuring participants' moralization of the issue and perceived threat from the opposing side, Study 3 allows us to test whether the “moderate as out-group” effect is contingent on these factors. Consistent with our account, we find that the effect occurs when the issue is seen as a matter of morality (vs. a matter of preference) and when the opposing side is perceived as a threat to oneself and/or close ones or to society at large, but it vanishes when either condition is absent.

Method

Participants

Nine hundred seventy-five Prolific workers from the United States completed the experiment in exchange for monetary compensation (\$0.35 for an expected 2-min completion time). As preregistered, participants who failed to correctly answer an attention-check question were not allowed to start the study, and participants who provided inconsistent ratings of the target (i.e., reporting that a target indicating “−3 = completely support Donald Trump” was not conservative, that a target indicating “3 = completely support Kamala Harris” was not liberal, that a target indicating “−2” was more conservative than one indicating “−3,” a target indicating “−1” was more conservative than one indicating “−2,” and so on for all targets) were excluded from the analyses. Our final sample consisted of 810 participants (390 identified as female, 404 identified as male, and 16 identified as nonbinary; $M_{\text{age}} = 39.7$ years, $SD_{\text{age}} = 12.5$ years).

Procedure

Participants first reported their political affiliation on a 5-point bipolar scale (−2 = *definitely conservative*, −1 = *mostly conservative*, 0 = *neither conservative nor liberal*, 1 = *mostly liberal*, 2 = *definitely liberal*). They then evaluated seven hypothetical individuals who expressed their position regarding support for one presidential candidate over the other in upcoming 2024 presidential elections. Each target expressed their position by indicating a different point on a 7-point scale (−3 = *completely support Donald Trump*, 0 = *equally support both*, 3 = *completely support Kamala*

Harris). Participants then judged these seven targets on a 5-point bipolar scale ($-2 = \text{target is definitely conservative}$, $2 = \text{target is definitely liberal}$).^{9,10} Participants also indicated whether they considered supporting one candidate over the other to be a matter of morality or a matter of political preference on a 5-point scale ($-2 = \text{entirely a matter of morality}$, $0 = \text{equally a matter of morality and of political preference}$, $2 = \text{entirely a matter of political preference}$). They next reported whether they considered the eventual election of the candidate representing the opposing political ideology¹¹ to be a threat (a) to themselves personally or someone they know and (b) to society at large. Both of these “threat” questions were asked on a 7-point scale ($1 = \text{not at all}$, $7 = \text{very much}$). All scales (except for the two “threat” scales) and the order of the positions were reversed for half of the participants. Finally, participants provided demographic information: gender (female, male, other) and age (free-response box).

Results and Discussion

Replicating the “moderate as out-group” effect, we found that conservative participants (i.e., those who indicated being “mostly conservative” and “definitely conservative”) judged the moderate target (i.e., the target indicating equal support for both candidates) as marginally liberal: $M = 0.04$, $SD = 0.33$; one-sample t test against 0: $t(178) = 1.80$, $p = .074$; see the red point for the “EQUALLY” position in Panel A of Figure 5, whereas liberal participants (i.e., those who indicated being “mostly liberal” and “definitely liberal”) judged the same target as conservative: $M = -0.21$, $SD = 0.50$; one-sample t test against 0: $t(406) = -8.36$, $p < .001$; see the blue point for the “EQUALLY” position in Panel A of Figure 5. As in our previous studies, the analysis considering the more granular positions revealed that “definitely liberal” participants judged the moderate target as more conservative ($M = -0.30$, $SD = 0.62$) than “mostly liberal” participants did: $M = -0.13$, $SD = 0.37$; $t(269.3) = -3.15$, $p = .002$; see the pink and turquoise points for the “EQUALLY” position in Panel B of Figure 5. However, although “definitely conservative” participants judged the moderate target as directionally more liberal ($M = 0.12$, $SD = 0.44$) than “mostly conservative” participants did ($M = 0.02$, $SD = 0.28$), this difference did not reach significance: $t(65.2) = 1.58$, $p = .120$; see the red and olive-green points for the “EQUALLY” position in Panel B of Figure 5.

Once again, conservative and liberal participants judged the targets differently across all positions (all t s > 2.16 , p s $< .031$; see the red and blue points in Panel A of Figure 5). As in previous studies, we created a “position type” variable (moderate vs. not moderate), this time contrasting judgments of the moderate target (i.e., the one equally supporting both candidates) with the average of the judgments of all other targets. Restricting the data set to participants who expressed a partisan position, moralized the issue, and perceived it as a threat ($N = 333$), we ran a non-preregistered linear model regressing judgment on position type (moderate vs. not moderate), participants’ ideology (conservative vs. liberal), and their interaction, clustering the SEs by participant. Predictors were effect coded ($-1/+1$). This model revealed a significant main effect of ideology, such that conservatives judged all targets as more liberal than liberals did and vice versa ($b = 0.17$, $t = 8.16$, $p < .001$), and no main effect of position type¹² ($b = 0.01$, $t = 0.83$, $p = .406$). Finally, we found a significant interaction showing that opposing partisans exhibited a greater judgment gap when evaluating the

moderate target than when evaluating the nonmoderate ones ($b = 0.04$, $t = 2.67$, $p = .008$; see Figure 6). In other words, like in previous studies, judgments of moderates are most susceptible to partisan position.

As noted earlier, the main goal of this study was to identify the conditions under which the “moderate as out-group” effect emerges. We predicted that the effect would be more likely to occur when participants (a) viewed support for one candidate over the other as a matter of morality rather than political preference and (b) perceived the opposing candidate’s election as a threat. To test these hypotheses, we restricted the dataset to participants who expressed a partisan position ($N = 586$). We recoded our dependent variable—the judgment of the moderate target ($-2 = \text{target is definitely conservative}$, $2 = \text{target is definitely liberal}$)—based on each participant’s own political orientation, such that negative values indicated that the target was categorized as belonging to the out-group and positive values indicated that the target was categorized as belonging to the in-group.

We first examined whether moralization predicted judgments of the moderate target. Regressing the recoded variable onto whether participants viewed support for one candidate over the other as a matter of morality or of political preference revealed that, as predicted, moralization increased the likelihood of categorizing moderates as out-group members: $b = 0.09$, $t(584) = 5.03$, $p < .001$; see Panel A of Figure 7. We next examined perceived threat, estimating two models—one regressing recoded judgment onto whether participants viewed the opposing candidate’s election as a threat to themselves or to someone they know, and one regressing recoded judgment onto whether participants viewed the opposing candidate’s election as a threat to society. As predicted, both measures predicted greater out-group categorization of the moderate target: $b_{\text{ThreatToSelf}} = -0.04$, $t(584) = -4.23$, $p < .001$; $b_{\text{ThreatToSociety}} = -0.06$, $t(584) = -4.67$, $p < .001$.¹³ Building beyond the preregistered analysis, because the two “threat” variables were highly correlated: $r = .80$, $t(584) = 32.69$, $p < .001$, we averaged them into a “general threat” variable and regressed the recoded judgment onto this new averaged-threat variable, which replicated that greater perceived threat increases out-group categorization of moderates: $b = -0.06$, $t(584) = -4.67$, $p < .001$; see Panel B of Figure 7.

Taken together, these analyses reveal that both moralization and perceived threat independently predict the tendency to categorize moderates as out-group members.

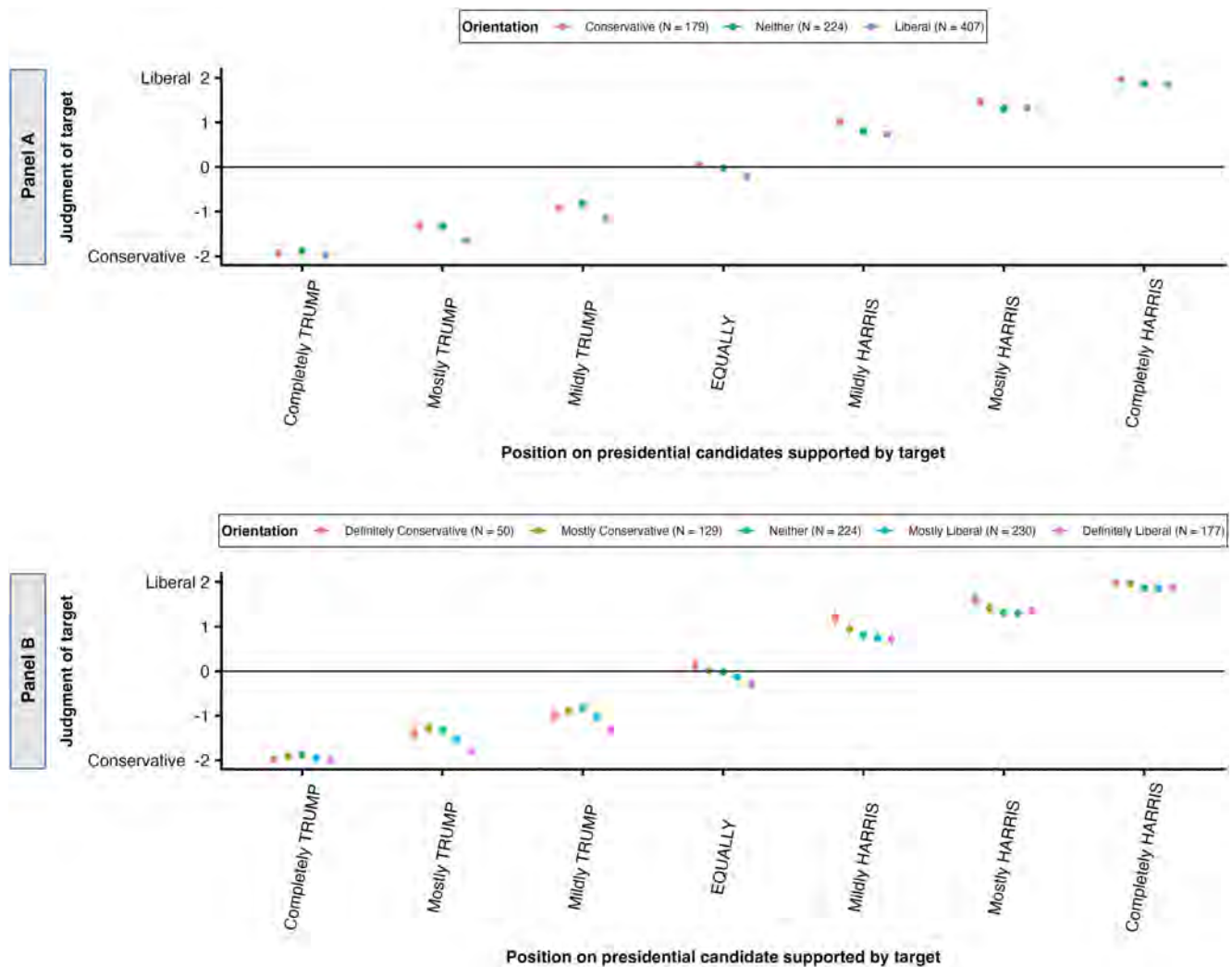
⁹ Note that we used a 5-point bipolar scale as our dependent variable to prevent participants from simply tracking the points indicated on the 7-point bipolar “support Trump versus Harris” scale (independent variable) onto the dependent variable. However, it was still straightforward to track our point of interest on the “support Trump versus Harris” scale (i.e., $0 = \text{equally support both}$) on the 5-point bipolar scale ($0 = \text{target is neither conservative nor liberal}$).

¹⁰ See ResearchBox for accessing the study materials.

¹¹ If they reported being liberal, they were asked about the eventual election of Donald Trump, and if they reported being conservative, they were asked about the eventual election of Kamala Harris. If they reported being neither conservative nor liberal, they were not asked this question.

¹² Unlike Studies 1 and 2, the positions in the “nonmoderate positions” level in Study 3 are balanced—three are conservative and three are liberal—which is why there is no main effect of position type.

¹³ Notably, two additional preregistered models—one regressing recoded judgment on moralization and threat to self, and one regressing recoded judgment on moralization and threat to society—revealed that all variables still significantly predicted out-group categorization when controlling for the other variables (all p s $< .0013$).

Figure 5*Study 3: “Moderate as Out-Group” Effect*

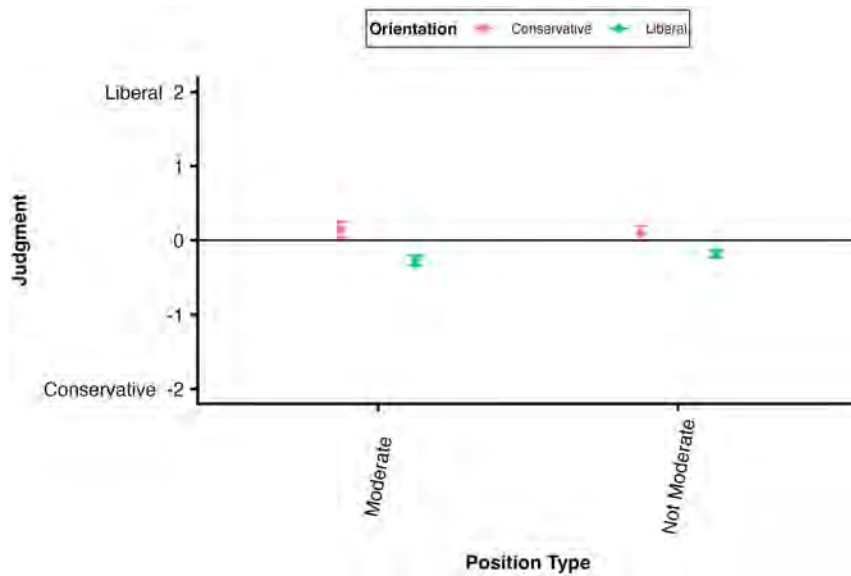
Note. The figure shows the results of Study 3. Panel A shows the judgment of the targets by three groups of participants (conservative vs. neither vs. liberal). Panel B shows the judgment of the targets by five groups (definitely conservative vs. mostly conservative vs. neither vs. mostly liberal vs. definitely liberal). Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

We next examined whether these conditions were individually sufficient to produce the “moderate as out-group” effect, or both were necessary. To do so, we conducted an exploratory (non-preregistered) analysis. As shown in Panel B of Figure 7, the relationship between threat and out-group categorization was not linear, concentrated at the higher end of the scale (ratings 5–7). Accordingly, we dichotomized the threat variable (1 = *high threat* for ratings 5.01–7 and 0 = *low threat* for ratings 1–5)¹⁴ and estimated separate models within each subsample. Specifically, we regressed recoded judgment (–2 = *definitely out-group*, 2 = *definitely in-group*) onto moralization, once for the “low threat” group ($N = 155$) and once for the “high threat” group ($N = 431$). Among participants reporting low threat perceptions, moralization did not predict judgment of the moderate target: $b = 0.01$, $t(153) = 0.52$, $p = .604$; see the left panel in Figure 8. Among participants reporting high threat perceptions, moralization significantly increased the likelihood

of categorizing the moderate target as an out-group member: $b = 0.09$, $t(429) = 3.99$, $p < .001$; see the right panel in Figure 8. Consistent with this pattern, a linear model regressing the recoded judgment onto moralization, dichotomized threat, and their interaction revealed a main effect of both moralization: $b = 0.09$, $t(582) = 4.57$, $p < .001$, and perceived threat: $b = 0.19$, $t(582) = 4.42$, $p < .001$, as well as a marginally significant interaction: $b = -0.08$, $t(582) = -1.96$, $p = .051$, showing that moralizing the issue has a larger effect when the threat level of the opposing side is perceived as high than when it is perceived as low. Together, these findings suggest that both moralization and

¹⁴ Note that the “threat” scale is the average of the “threat to self” and “threat to society” measures—for this reason, we have decimal “threat” ratings.

Figure 6
Study 3: Judgment Gap for Moderate and Nonmoderate Targets



Note. The figure shows the results of Study 3. Judgment of targets by participants' political ideology (conservative vs. liberal) and position type (moderate vs. not moderate). Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

perceived threat are necessary for the “moderate as out-group” effect to emerge, and neither is sufficient on its own.

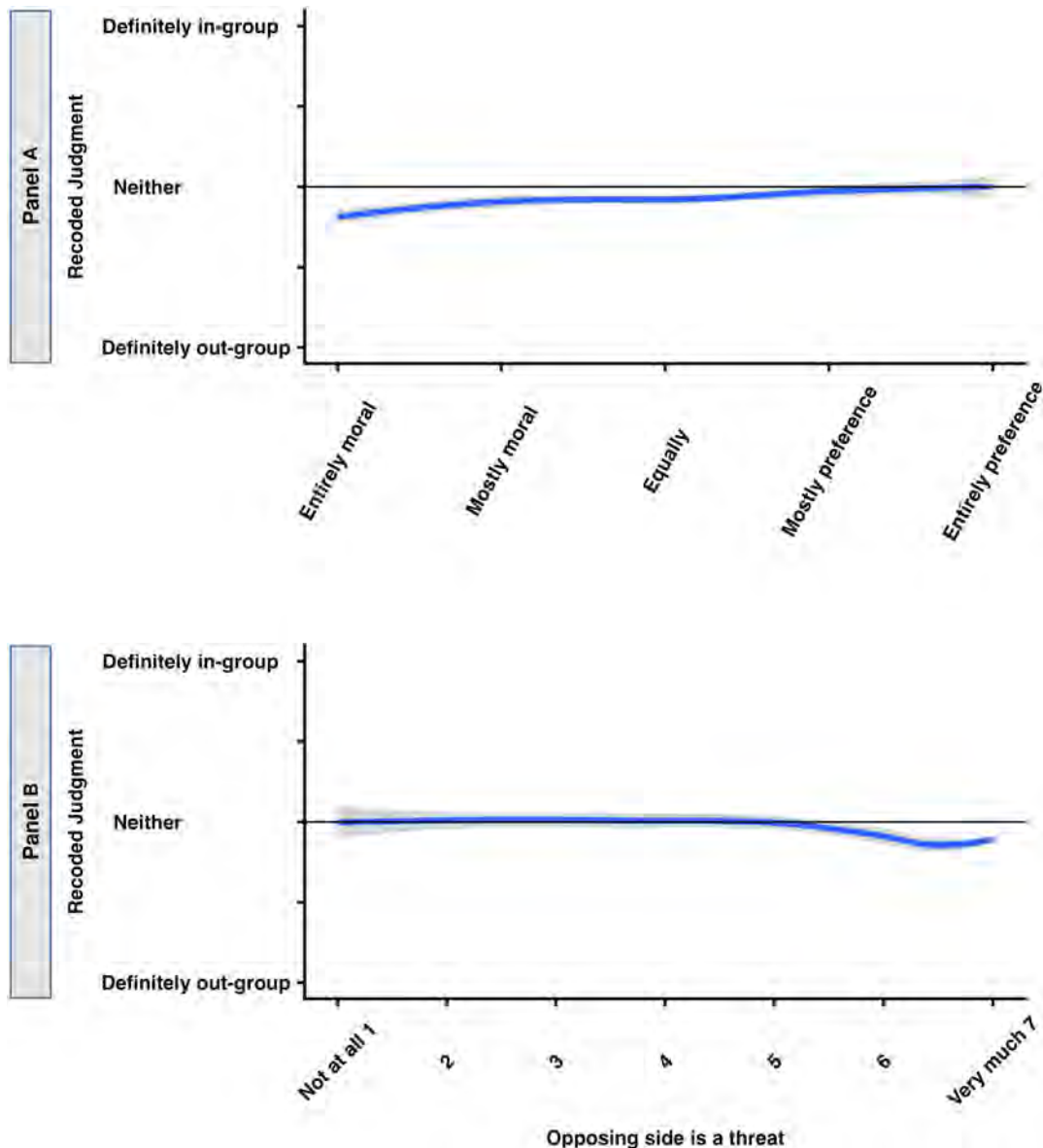
Importantly, these results also bear on the role of naïve realism. Naïve realism predicts that when individuals view their own positions as objectively correct, disagreement is interpreted as bias. Because moralized views are especially likely to be perceived as objective truth (e.g., Skitka et al., 2005), naïve realism would predict that moralized disagreement should produce biased categorizations regardless of perceived threat. Our results do not support this prediction. Participants who moralized the issue but did not perceive the opposing side as threatening did not exhibit the “moderate as out-group” effect, suggesting that epistemic certainty alone is insufficient to produce these category boundary shifts.

Finally, because ideological extremity, moralization, and perceived threat are conceptually related, we examined their interrelations using Pearson correlations. The three variables were modestly correlated. Greater moralization of the issue (i.e., viewing it more as a matter of morality than political preference) was associated with greater perceived threat from the opposing side ($r = -.23, p < .001$). More extreme participants were also more likely to view the issue as moral rather than a matter of political preference ($r = -.18, p < .001$) and to perceive the opposing side as more threatening ($r = .33, p < .001$). These relationships may help explain why participants with more extreme views exhibit a larger “moderate as out-group” effect.

Study 3 identifies the conditions under which the “moderate as out-group” effect is most likely to emerge—namely, when socio-political issues are both moralized and the opposing side is perceived as threatening. This finding provides potential insight into the psychological mechanism driving the effect. Specifically, it suggests that the effect hinges on the moral and threatening nature of the issue.

In Studies 4 and 5, we directly test the psychological mechanism underlying the “moderate as out-group” effect. We hypothesize that this effect arises because people perceive that the moderate target agrees with—or simply does not condemn—the immoral and threatening out-group ideology (absence of the “out-group hate” motive), rather than disagrees with their own ideology (absence of the “in-group love” motive). For example, consider a moderate policy on abortion for nonmedical reasons, such as “Abortion allowed up to 8 weeks from conception.” Partisan individuals might perceive a person supporting this policy as aligning with the pro-choice ideology—because it allows women to choose, indicating a value for personal choice—and/or with the pro-life ideology—because the time window for choice is narrow, indicating a value for the life of the fetus. This ambiguity could yield either a “pro-both sides” interpretation (the moderate also supports the other side) or a “pro-neither side” interpretation (the moderate does not sufficiently support my side). We predict that partisans tend to adopt the former (pro-both sides) interpretation, reflecting the absence of an “out-group hate” motive, which elicits strong aversion (consistent with Gershon & Fridman, 2022) and drives the “moderate as out-group” effect. We test this hypothesis in two different ways. In Study 4, we manipulate whether the moderate stance concerns an out-group target (implying absence of “out-group hate”) or an in-group target (implying absence of “in-group love”). In Study 5, we manipulate the framing of a moderate position as either pro-both sides (implying absence of “out-group hate”) or pro-neither side (implying absence of “in-group love”).

In Study 3, liberal participants perceived supporting the two presidential candidates as a matter of morality: $M = -0.27, SD = 1.06$; one-sample t test against 0: $t(406) = -5.15, p < .001$, and the opposing candidate as highly threatening: $M = 6.11, SD = 1.30$;

Figure 7*Study 3: Necessary Conditions for the “Moderate as Out-Group” Effect*

Note. The figure shows the results of Study 3. Panel A shows the trendline (a locally estimated scatterplot smoothing regression line) displaying the relationship between judging a moderate target as in-group/out-group and viewing the issue as a matter of morality versus political preference. Panel B shows the trendline displaying the relationship between judging a moderate target as in-group/out-group and perceiving the opposing side as a (general) threat. Gray-shaded areas indicate the 95% confidence intervals. See the online article for the color version of this figure.

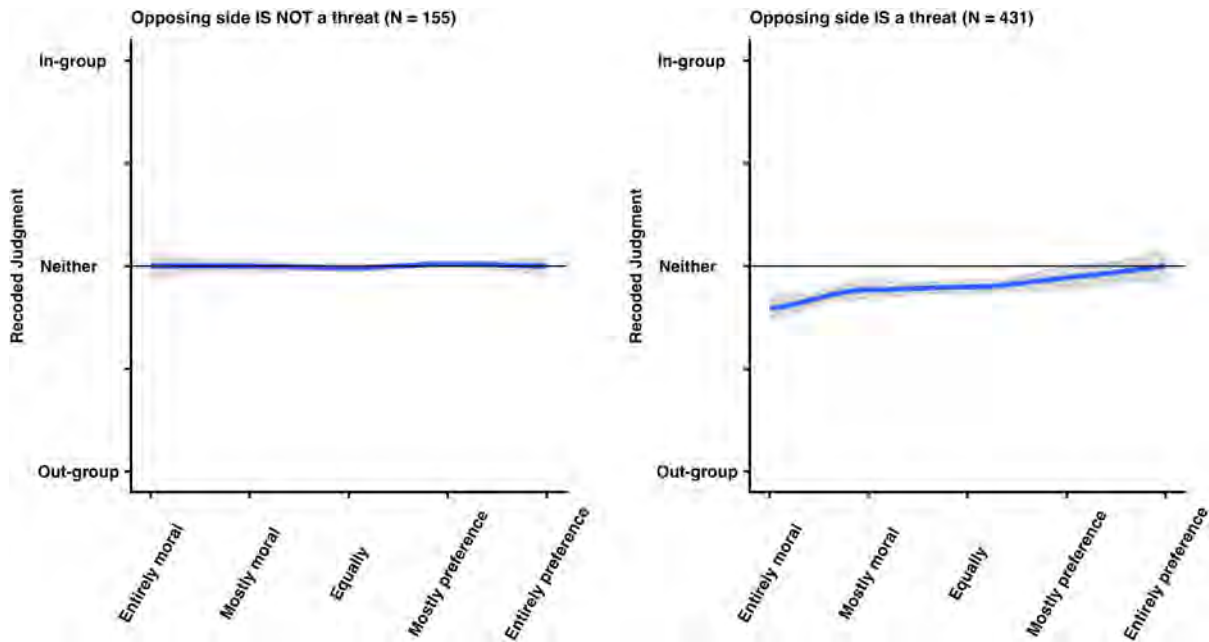
one-sample t test against 5¹⁵: $t(406) = 17.19, p < .001$. They also rated the issue as more moralized: $t(339.39) = -5.31, p < .001$, and more threatening: $t(253.76) = 6.26, p < .001$, than conservatives did. For these reasons, in Study 4, we recruited only liberal participants and manipulated positions on the two candidates across conditions. This design allowed us to manipulate absence of “out-group hate” versus absence of “in-group love.” Holding a moderate position on Donald Trump (out-group to liberals) implies only absence of

out-group hate, whereas holding a moderate position on Kamala Harris (in-group to liberals) implies only absence of in-group love.

¹⁵ Although the midpoint of the scale was 4, we tested the mean against 5 for two reasons: (a) 5 was the threshold used to dichotomize the threat variable, and (b) this provides a more conservative test of threat perceptions, as participants needed to report above-midpoint threat to differ significantly from the test value.

Figure 8

Study 3: Necessary but Not Sufficient Conditions for the “Moderate as Out-Group” Effect



Note. The figure shows the results of Study 3. Trendlines (locally estimated scatterplot smoothing regression lines) displaying the relationship between the judgment of a moderate target as belonging to the in-group/out-group and viewing the sociopolitical issue at hand as a matter of morality versus political preference when the opposing side is not perceived as a threat and when it is. Gray-shaded areas indicate the 95% confidence intervals. See the online article for the color version of this figure.

Study 4: The “Moderate as Out-Group” Effect Is Driven by Absence of Out-Group Hate Rather Than by Absence of In-Group Love

Study 4 tests the hypothesis that the “moderate as out-group” effect is driven by the perceived absence of an “out-group hate” motive rather than by the absence of an “in-group love” motive. This study employs a different experimental paradigm from the previous ones: We recruited only liberal participants, who were randomly assigned to evaluate seven targets expressing varying levels of support for either Donald Trump (out-group to liberals) or Kamala Harris (in-group to liberals). Because out-group hate tends to outweigh in-group love in intergroup conflicts (e.g., Gershon & Fridman, 2022; Halevy et al., 2008; Rathje et al., 2021), we predicted that liberals would exhibit the “moderate as out-group” effect when judging a target expressing a moderate stance toward Donald Trump (implying absence of out-group hate motive), but not when judging a target expressing a moderate stance toward Kamala Harris (implying absence of in-group love motive).

Method

Participants

Nine hundred four Prolific workers from the United States who had indicated being Democrats on the platform completed the experiment in exchange for monetary compensation (\$0.35 for an expected 2-min completion time). As preregistered, participants who failed an attention-check question or indicated a political affiliation

other than “definitely liberal” or “mostly liberal” on a 5-point bipolar scale (ranging from $-2 = \text{definitely conservative}$ to $2 = \text{definitely liberal}$) were not allowed to start the study.¹⁶ Additionally, participants who provided inconsistent ratings of the target—using the same specifications as in Study 3—were excluded from the analyses. The final sample consisted of 796 participants (384 identified as female, 396 identified as male, and 16 identified as nonbinary; $M_{\text{age}} = 43.0$ years, $SD_{\text{age}} = 14.0$ years).

Procedure

As in Study 3, participants first reported their political affiliation on a 5-point bipolar scale ($-2 = \text{definitely conservative}$, $-1 = \text{mostly conservative}$, $0 = \text{neither conservative nor liberal}$, $1 = \text{mostly liberal}$, $2 = \text{definitely liberal}$) and then evaluated seven hypothetical individuals. This time, however, the hypothetical targets expressed their position on either Donald Trump’s politics—indicating each point of a 7-point scale ($-3 = \text{completely support Trump’s politics}$, $0 = \text{mixed opinions about Trump’s politics}$, $3 = \text{completely oppose Trump’s politics}$)—or Kamala Harris—indicating each point of a 7-point scale ($-3 = \text{completely oppose Harris’ politics}$, $0 = \text{mixed}$

¹⁶ Note that we preregistered recruiting 1,000 liberal participants. The reason only 904 participants completed the study is that—despite recruiting exclusively from Prolific workers who indicated being Democrats on the platform—at the beginning of the study, 103 individuals indicated that they were either conservative (26) or neither conservative nor liberal (77) and were therefore not allowed to start the study, as preregistered. Of these, 96 nevertheless submitted the study for payment on Prolific, and we compensated them, which explains the observed discrepancy.

opinions about Harris' politics, 3 = *completely support Harris' politics*)—depending on the condition to which participants were randomly assigned. Participants then judged the seven assigned targets on a 5-point bipolar scale ($-2 = \text{target is definitely conservative}$, $2 = \text{target is definitely liberal}$).^{17,18} All scales and the order of the positions were reversed for half of the participants. Finally, participants indicated their gender (female, male, other) and age (free-response box).

Results and Discussion

As predicted, and replicating the results from previous studies, when liberals evaluated the target expressing a moderate position on Donald Trump's politics, we observe a robust "moderate as out-group" effect. Liberal participants categorized the target indicating "mixed opinions about Trump's politics" as conservative: that is, rated significantly lower than 0; $M = -0.50$, $SD = 0.68$; one-sample t test against 0: $t(408) = -15.00$, $p < .001$; see the red point for the "MIXED opinions" position in Panel A of Figure 9. However—also as predicted—liberals categorized the target indicating "mixed opinions about Harris' politics" as moderate: that is, not significantly different from 0; $M = 0.03$, $SD = 0.50$; one-sample t test against 0: $t(386) = 1.21$, $p = .226$; see the blue point for the "MIXED opinions" position in Panel A of Figure 9. A two-sample t test confirmed that liberals' judgments of the targets expressing moderate positions about Trump's versus Harris' politics differed significantly: $t(752.86) = -12.65$, $p < .001$; see the red and blue points for the "MIXED opinions" position in Panel A of Figure 9. In addition, and as in our previous studies, "definitely liberal" participants judged the target expressing a moderate position on Trump's politics as more conservative ($M = -0.62$, $SD = 0.72$) than "mostly liberal" participants did: $M = -0.38$, $SD = 0.61$; $t(403.69) = -3.70$, $p < .001$; see the red points for the "MIXED opinions" position in Panel B of Figure 9. However—further underscoring the robustness of this moderation—"definitely liberal" and "mostly liberal" participants judged the target expressing a moderate position on Harris' politics virtually identically: $M_{\text{DefinitelyLiberal}} = 0.03$, $SD = 0.59$; $M_{\text{MostlyLiberal}} = 0.03$, $SD = 0.40$; $t(350.95) = -0.03$, $p = .972$; see the blue points for the "MIXED opinions" position in Panel B of Figure 9.

Interestingly, liberal participants judged the targets expressing positions on Trump as more conservative than those expressing positions on Harris across the entire range of positions (all $ps < .001$), except for the targets indicating the "completely liberal" positions: that is, 3 = *completely oppose Trump's politics* and 3 = *completely support Harris' politics*; $t(793.93) = -0.74$, $p = .462$. This difference in judgment, however, was more pronounced for the moderate positions. In fact, a non-preregistered linear model regressing judgment on position type (moderate vs. not moderate),¹⁹ condition (positions on Trump vs. Harris), and their interaction, clustering the SEs by participant (predictors were effect coded $-1/+1$), revealed a significant main effect of position type, such that moderate targets were judged as more conservative than the aggregated nonmoderate targets ($b = -0.05$, $t = -5.99$, $p < .001$), and a significant main effect of condition, such that targets expressing positions on Trump's politics were judged as more conservative than those expressing positions on Harris' politics ($b = 0.18$, $t = 13.20$, $p < .001$). Finally, there was a significant interaction showing that liberals judged the targets expressing the moderate positions on Trump and Harris more differently than the targets

expressing the nonmoderate positions on Trump and Harris ($b = 0.08$, $t = 9.72$, $p < .001$; see Figure 10).

Study 4 provides evidence for the psychological mechanism underlying the "moderate as out-group" effect by demonstrating that the effect is driven by the perception that moderates lack the "out-group hate" motive rather than lack the "in-group love" motive.

Study 5: The "Moderate as Out-Group" Effect Occurs When Moderate Positions Are Framed as Pro-Both Sides, but Not as Pro-Neither Side

Study 5 tests the same hypothesis as Study 4—that the "moderate as out-group" effect is driven by partisans' perception that moderates agree with the opposing ideology (absence of out-group hate), rather than disagree with their own (absence of in-group love)—but does so in a different way. Here, we manipulate the framing of a moderate position as either pro-both sides or pro-neither side. Consistent with the findings from Study 4, we find that the effect occurs when the moderate position is framed as pro-both sides (implying absence of out-group hate) and it is attenuated when it is framed as pro-neither side (implying absence of in-group love).

The study was conducted on November 14, 2023—just 5 weeks after Hamas' October 7, 2023, attack on Israelis—and concerns positions on the conflict.

Method

Participants

Seven hundred sixty-four Prolific workers from the United States completed the experiment in exchange for monetary compensation (\$0.30 for an estimated 2-min completion time). As preregistered, participants who failed to correctly answer an attention-check question were not allowed to start the study. Additionally, participants who reported not being familiar with the recent events in the Middle East (i.e., those who indicate "1" or "2" on a 5-point familiarity-with-the-conflict scale; 1 = *not at all familiar*, 5 = *very familiar*) and participants who provided inconsistent ratings of Jamie (i.e., reporting Jamie supporting the pro-Israel position to be pro-Palestine, Jamie supporting the pro-Palestine position to be pro-Israel, Jamie supporting the moderate positions to be more pro-Israel than when supporting the pro-Israel position, or Jamie supporting the moderate positions to be more pro-Palestine than when supporting the pro-Palestine position) were excluded from the analyses. Specifically, 134 participants were excluded for unfamiliarity with the conflict, and 155 were excluded for providing inconsistent ratings of Jamie. The final sample consisted of 475 participants (209 identified as female, 262 identified as male, and four identified as nonbinary; $M_{\text{age}} = 41.0$ years, $SD_{\text{age}} = 13.8$ years).

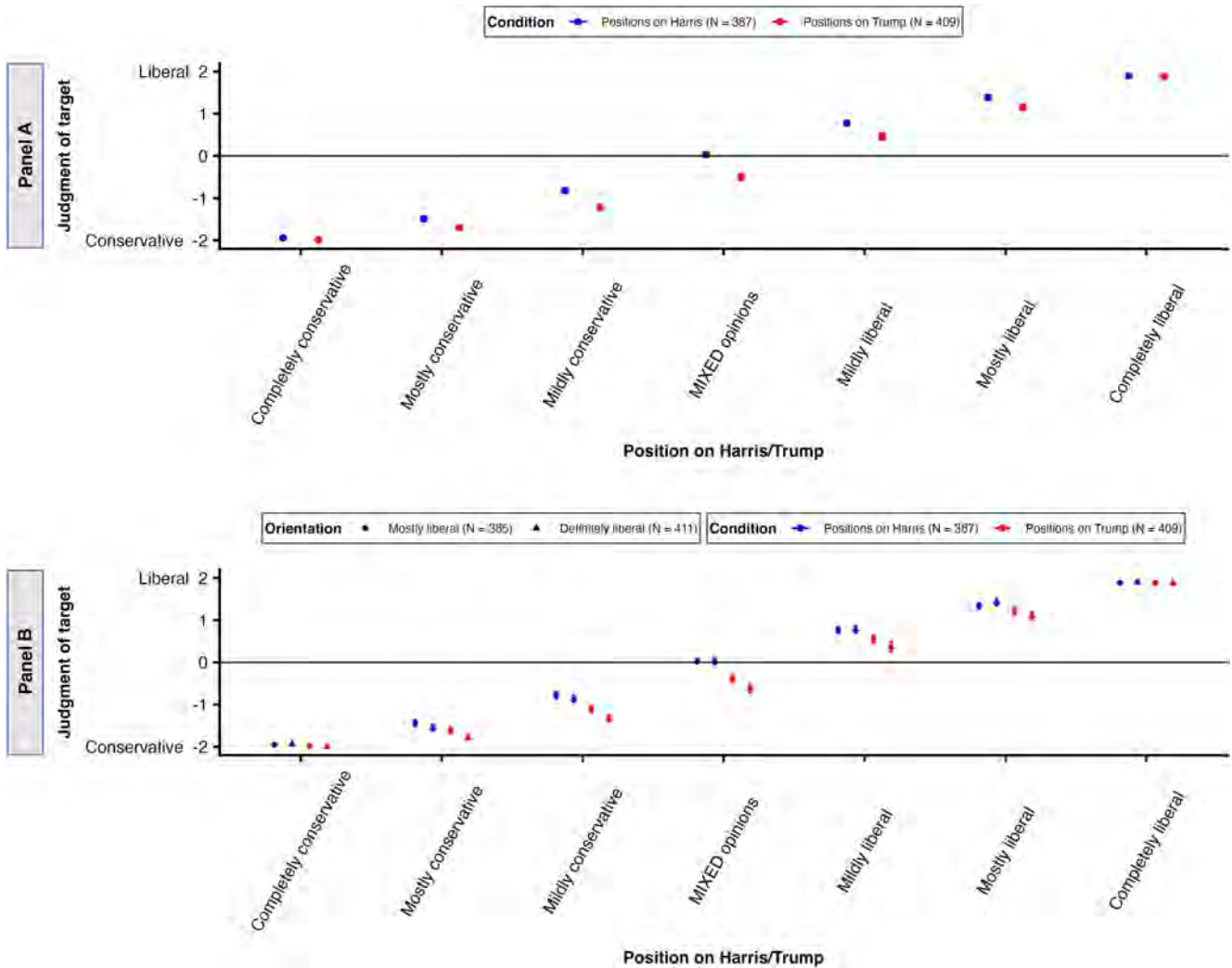
¹⁷ As in Study 3, we used a 5-point bipolar scale as our dependent variable to prevent participants from simply tracking the points indicated on the 7-point bipolar "support versus oppose Trump/Harris" scale (independent variable) onto the dependent variable.

¹⁸ See ResearchBox for accessing the study materials.

¹⁹ As in the previous studies, we created a "position type" variable (moderate vs. not moderate), contrasting the average of the judgments of the moderate targets (i.e., indicating having "mixed opinions" on Trump and Harris' politics) and the average of the nonmoderate targets (i.e., indicating any other point of the 7-point scales).

Figure 9

Study 4: Moderation of the “Moderate as Out-Group” Effect



Note. The figure shows the results of Study 4. Panel A shows the judgments of the targets by condition (positions on Harris vs. Trump). Panel B shows the judgments of the targets by condition (positions on Harris vs. Trump) and by participants' extremity (mostly liberal vs. definitely liberal). Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

Procedure

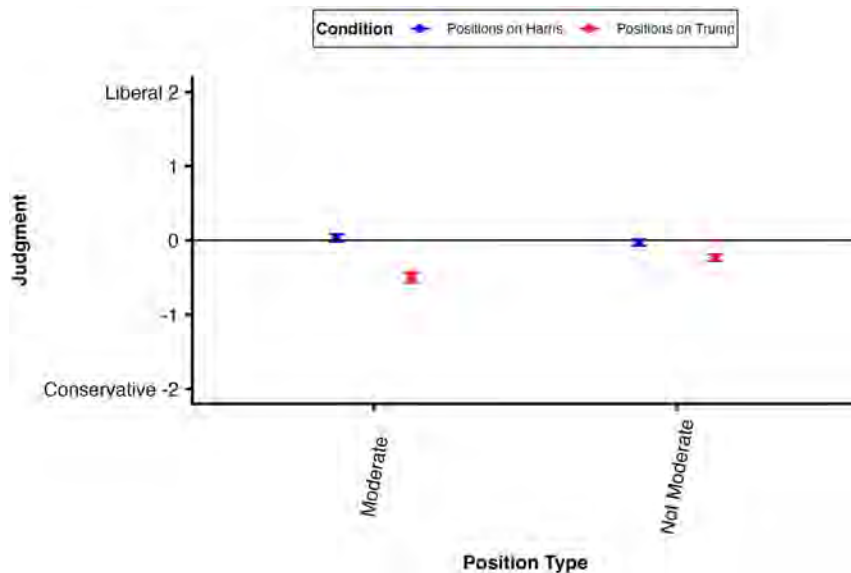
Participants first reported their familiarity with Hamas's October 7, 2023, attack on Israel and Israel's response using a 5-point unipolar scale (1 = *not at all familiar*, 2 = *barely familiar*, 3 = *somewhat familiar*, 4 = *mostly familiar*, 5 = *very familiar*). They then reported their position on the Israel–Palestine conflict on a 5-point bipolar scale (−2 = *definitely pro-Israel*, −1 = *mostly pro-Israel*, 0 = *neutral*, 1 = *mostly pro-Palestine*, 2 = *definitely pro-Palestine*) and evaluated three positions on the conflict—one definitely pro-Israel, one definitely pro-Palestine, and a moderate one. Next, participants judged a hypothetical person, Jamie, holding each position on a 7-point bipolar scale (−3 = *Jamie is definitely pro-Israel*, 3 = *Jamie is definitely pro-Palestine*). Participants were randomly assigned to consider one of two moderate positions: either “Both Hamas's October 7th attack on Israelis and Israel's subsequent siege of Gaza are acceptable” (“pro-both sides”

condition) or “Neither Hamas's October 7th attack on Israelis nor Israel's subsequent siege of Gaza are acceptable” (“pro-neither side” condition). We refer to these conditions as “pro-both sides” and “pro-neither side,” but participants were shown the full description of a moderate position (as reported above). Finally, participants reported their political affiliation on a 5-point bipolar scale (−2 = *extremely conservative*, 2 = *extremely liberal*) and provided demographic information: gender (female, male, other) and age (free-response box). As in previous studies, all scales and the order of the positions were reversed for half of the participants.

Results and Discussion

We first analyzed the judgments of three groups of participants: pro-Israel (those who indicated being “mostly pro-Israel” or “definitely pro-Israel”), pro-Palestine (those who indicated being “mostly

Figure 10
Study 4: Judgment Gap for Moderate and Nonmoderate Positions



Note. The figure shows the results of Study 4. Judgment of targets by condition (positions on Trump vs. Harris) and position type (moderate vs. not moderate). Error bars represent 95% confidence intervals. See the online article for the color version of this figure.

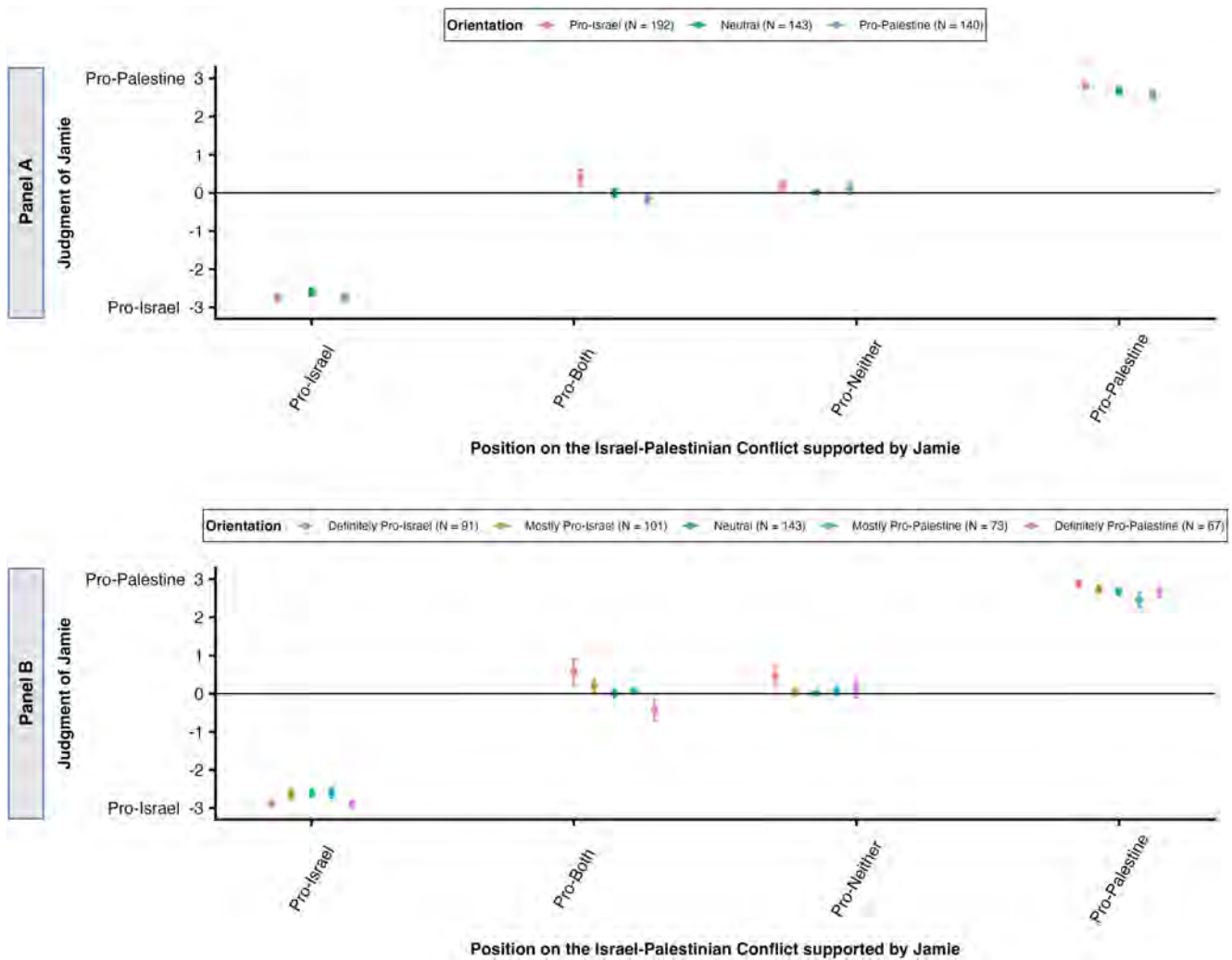
pro-Palestine” or “definitely pro-Palestine”), and neither pro-Israel nor pro-Palestine. As predicted, when the moderate position was expressed as pro-both sides, we observed the “moderate as out-group” effect. Pro-Israel participants judged Jamie as pro-Palestine: that is, rated significantly greater than 0; $M = 0.39$, $SD = 1.00$; one-sample t test against 0: $t(93) = 3.83$, $p < .001$; see the red point for “pro-both” position in Panel A of Figure 11, whereas pro-Palestine participants judged Jamie as pro-Israel: that is, rated significantly lower than 0; $M = -0.15$, $SD = 0.64$; one-sample t test against 0: $t(78) = -2.10$, $p = .039$; see the blue point for “pro-both” position in Panel A of Figure 11. As in our previous studies, the effect was more pronounced among participants with more extreme views. “Definitely pro-Israel” participants judged Jamie supporting the “pro-both sides” position as marginally more pro-Palestine ($M = 0.56$, $SD = 1.23$) than “mostly pro-Israel” participants did: $M = 0.20$, $SD = 0.59$; $t(72.5) = 1.82$, $p = .074$; see the red and olive-green points for the “pro-both sides” position in Panel B of Figure 11, and “definitely pro-Palestine” participants judged Jamie supporting the same position as more pro-Israel ($M = -0.44$, $SD = 0.86$) than “mostly pro-Palestine” participants did: $M = 0.07$, $SD = 0.25$; $t(37.3) = -3.34$, $p = .002$; see the pink and turquoise points for the “pro-both sides” position in Panel B of Figure 11.

Importantly, the “moderate as out-group” effect was attenuated when the moderate position was expressed as pro-neither side. The difference in judgment between pro-Israel and pro-Palestine participants that was observed when the moderate position was framed as pro-both sides: $M_{\text{Pro-Israel}} = 0.39$, $SD = 1.00$; $M_{\text{Pro-Palestine}} = -0.15$, $SD = 0.64$; $t(160.8) = 4.34$, $p < .001$, disappeared when it was framed as pro-neither side: $M_{\text{Pro-Israel}} = 0.20$, $SD = 0.70$; $M_{\text{Pro-Palestine}} = 0.11$, $SD = 0.58$; $t(144.9) = 0.87$, $p = .386$. To further investigate the predicted moderation, we restricted the data only to the judgments of Jamie supporting the moderate positions made by partisan participants ($N = 332$) and

conducted a 2 (conflict orientation: pro-Israel vs. pro-Palestine) $\times$ 2 (moderate-position framing: pro-both vs. pro-neither side) analysis of variance on judgment of Jamie, which revealed a main effect of conflict orientation, such that pro-Israel participants judged Jamie as more pro-Palestine than pro-Palestine participants did: $F(1, 328) = 13.75$, $p < .001$, no main effect of moderate-position framing: $F(1, 328) = 0.20$, $p = .653$, and—confirming the presence of a moderation—a significant interaction, such that the judgments of opposing partisans differed more when Jamie supported the pro-both position than the pro-neither one: $F(1, 328) = 7.10$, $p = .008$; see the red and blue points for “pro-both” and “pro-neither” positions in Panel A of Figure 11. This result replicates if we restrict the data set only to participants with more extreme views of the conflict (i.e., “definitely pro-Israel/pro-Palestine”), despite a smaller sample size ($N = 158$)—2 $\times$ 2 analysis of variance revealed a significant main effect of conflict orientation: $F(1, 154) = 16.05$, $p < .001$; no main effect of moderate-position framing: $F(1, 154) = 2.15$, $p = .145$; a significant interaction: $F(1, 154) = 4.92$, $p = .028$ (see the red and pink points for “pro-both” and “pro-neither” positions in Panel B of Figure 11). Among these extreme participants, the judgment of “definitely pro-Israel” and “definitely pro-Palestine” individuals differed when the moderate position was framed as pro-both sides: $M_{\text{Definitely Pro-Israel}} = 0.56$, $SD = 1.23$; $M_{\text{Definitely Pro-Palestine}} = -0.44$, $SD = 0.86$; $t(81.9) = 4.39$, $p < .001$, but not when it was framed as pro-neither side: $M_{\text{Definitely Pro-Israel}} = 0.44$, $SD = 0.95$; $M_{\text{Definitely Pro-Palestine}} = 0.15$, $SD = 0.76$; $t(72.0) = 1.45$, $p = .151$.

As in the previous studies, we created a “position type” variable (moderate vs. not moderate) by contrasting judgments of Jamie supporting the pro-both position²⁰ with the average of the judgments

²⁰ Note that we did not include the observations from the “pro-neither” condition, because the effect is moderated.

Figure 11*Study 5: Moderation of the “Moderate as Out-Group” Effect*

Note. The figure shows the results of Study 5. Panel A shows the judgments of Jamie by three groups of participants (pro-Israel vs. neither vs. pro-Palestine). Panel B shows the judgments of Jamie by five groups (definitely pro-Israel vs. mostly pro-Israel vs. neither vs. mostly pro-Palestine vs. definitely pro-Palestine). Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

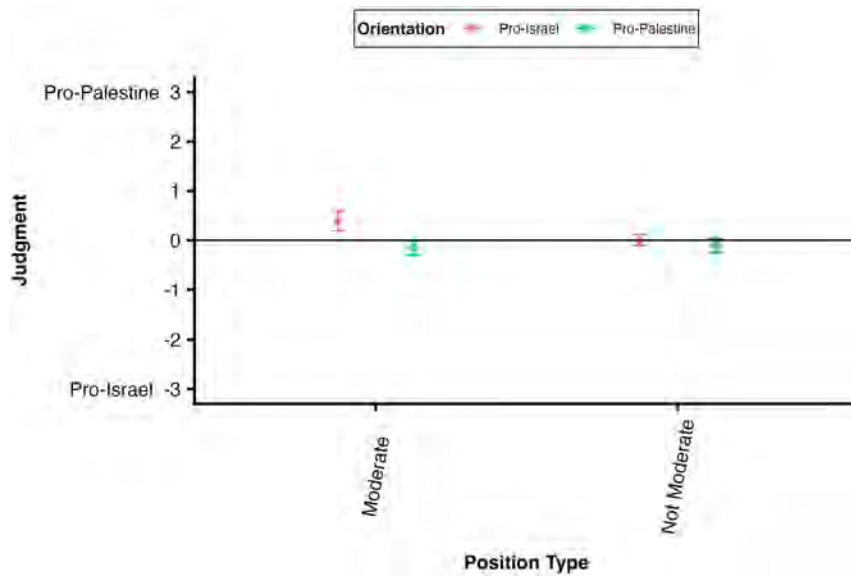
of Jamie supporting the pro-Israel and pro-Palestine positions. After restricting the data set to participants in the pro-both condition who expressed a partisan position ($N = 173$), we conducted a non-preregistered linear model regressing judgment on position type (moderate vs. not moderate), participants' conflict orientation (pro-Israel vs. pro-Palestine), and their interaction, clustering the SEs by participant. Predictors were effect coded ($-1/+1$). This revealed a significant main effect of orientation, such that pro-Israel participants judged Jamie as more pro-Palestine across all positions than pro-Palestine participants did ($b = 0.16$, $t = 4.48$, $p < .001$). It also revealed a significant main effect of position type, such that Jamie supporting the moderate pro-both position was judged as more pro-Palestine than when supporting the nonmoderate positions and vice versa ($b = 0.09$, $t = 2.72$, $p = .007$). Finally, it showed a significant interaction, indicating that opposing partisans exhibited a larger judgment gap when evaluating Jamie supporting the moderate

pro-both position than the nonmoderate ones ($b = 0.11$, $t = 3.42$, $p < .001$; see Figure 12).

Complementing the results of Study 4, Study 5 offers further evidence that the “moderate as out-group” effect is driven by the perception that moderates agree with—or even just fail to condemn—the out-group (absence of an “out-group hate” motive), rather than by the perception that moderates disagree with the in-group (absence of an “in-group love” motive). This study also shows that, despite dual agreement, pro-both positions are nonetheless categorized as belonging to the out-group ideology, showing that agreement with the out-group is more diagnostic in categorization judgments than agreement with the in-group. In other words, the lack of out-group derogation is treated as a stronger signal of group membership than in-group endorsement.

Beyond these theoretical insights, Study 5 also points to practical ways of expressing moderate positions without triggering backlash from all sides. Specifically, it suggests that moderation is more

Figure 12
Study 5: Judgment Gap for Moderate and Nonmoderate Positions



Note. The figure shows the results of Study 5. Judgment of Jamie by participants' orientation on the conflict (pro-Israel vs. pro-Palestine) and position type (moderate vs. not moderate). Error bars represent the 95% confidence intervals. See the online article for the color version of this figure.

acceptable when framed as disagreement with both sides, rather than as agreement with both.

General Discussion

In this work, we examined whether individuals with partisan views on a sociopolitical issue categorize moderate people and policies as moderate, as belonging to their in-group, or as belonging to the out-group. Across five preregistered studies, we consistently found that, under specific circumstances, partisans tend to categorize moderates as out-group members—a phenomenon we term the “moderate as out-group” effect. We showed that the effect occurs when an issue is both moralized (and therefore polarizing) and the opposing side is perceived as threatening. We further demonstrated that more extreme partisans are more likely to moralize issues and to perceive the opposing side as threatening, which helps explain why, across all studies, they exhibited the “moderate as out-group” effect more strongly than less extreme partisans. While we consistently observed that partisan extremity exacerbated the effect, only in some studies did the effect extend to less extreme partisans. We suspect that this depends on the extent to which the issue or target is moralized and perceived as threatening. When those perceptions are concentrated among the highly extreme, the effect is largely limited to them, whereas when they are more broadly shared, the effect extends to less extreme partisans as well, although to a lesser degree. Several factors may influence these patterns, including whether participants are evaluating abstract positions or specific individuals and the degree to which the issue itself is moralized or experienced as threatening. In this sense, the variation observed across studies is consistent with our theoretical framework and likely reflects

differences in how strongly these processes are activated across contexts and groups. Finally, we identified the underlying psychological mechanism of the “moderate as out-group” effect. It is driven by the perception that moderates are receptive to—or fail to condemn—the immoral and threatening out-group (absence of an “out-group hate” motive), rather than by the perception that they disagree with their own side (absence of an “in-group love” motive).

Ruling Out Two Alternative Explanations

We considered and ruled out two alternative explanations for the “moderate as out-group” effect. First, we considered that partisans may misperceive the distribution of positions on the issue in the general population, overestimating the prevalence of their own views due to living in ideologically homogeneous social networks. In other words, they might believe that the average opinion held by the population is closer to their own side than it actually is. This skewed perception of the distribution of opinions on the matter could lead them to infer that moderates lie on the tail of such biased distribution and therefore belong to the out-group. Second, we considered that partisans might perceive their own positions as more moderate than they actually are, thereby shifting their internal definition of “moderate” toward their in-group. Truly moderate positions might then be perceived as crossing the partisan threshold, into the out-group territory.

Importantly, Studies 4 and 5 undermine both of these alternative explanations. Neither biased distribution perceptions nor biased definitions of moderation can account for the attenuations of the effect observed when a target is moderate on an in-group topic or person, or when a moderate position is framed as pro-neither side.

Such manipulations should not affect people's perceived distributions of positions in the population or their self-concept as moderate versus extreme, yet they produced robust attenuations of the effect (see Studies 4 and 5).

Contributions and Implications of the “Moderate as Out-Group” Effect

Our findings contribute to several ongoing theoretical debates in social psychology and political science. First, they extend group categorization and polarization research by showing that moderates are treated not as boundary spanners but as out-group members when issues are moralized and threatening. Second, they advance attribution research by demonstrating a motivational asymmetry in group-membership judgments: Perceived agreement with the out-group is weighted more heavily than agreement with the in-group. Third, they complement—and extend beyond—accounts of false polarization (e.g., Keltner & Robinson, 1996; Lees & Cikara, 2021) by showing that the perceptual distortion not only concerns extremity of the out-group, but disproportionately affects moderates, whose categorization also crosses category boundaries in ways that the categorization of nonmoderates does not.

Our findings also have implications for the dynamics of political polarization and social cohesion. Sociopolitical issues are becoming increasingly moralized—consistent with a process described as “concept creep” (Haslam, 2016). This suggests that the “moderate as out-group” effect is likely to become both more frequent and more relevant over time. One potential implication is that polarization may be self-reinforcing: If moderates are rejected by both sides, they may feel compelled to abandon moderation and align with a partisan camp, further deepening divides. The “moderate as out-group” effect may also help explain the decline of nonpartisan news outlets in the United States (e.g., Finkel et al., 2020). Over time, even traditionally centrist outlets such as CNN have adopted a clear partisan lean, joining Fox News on the right and MSNBC on the left. Recent research was able to measure the airtime of conservative and liberal guest speakers on these three major outlets and found that, from 2010 to 2021, all of them exhibited a marked partisan bias: Fox News on the conservative side, and MSNBC and CNN on the liberal side (with CNN's liberal bias being almost as large as MSNBC's; Kim et al., 2022).

Our results may also offer new insight into the “hostile media” effect (e.g., Vallone et al., 1985), whereby partisans perceive identical news coverage of controversial issues as biased against their own side (Feldman, 2014). Vallone et al. (1985) proposed three drivers for the “hostile media” effect: (a) naïve realism (or “different standards”), which posits that, because partisans believe their side's claims to be more accurate than the other side's, they believe media coverage that reports an equal number of claims from both parts to be inappropriate and unfairly biased against their side; (b) selective categorization, according to which partisans on each side attend to different parts of the information they receive, typically attending to the hostile ones more; and (c) selective recall, according to which partisans remember hostile pieces of information more. Our results suggest a different take on the psychological forces driving the “hostile media” effect, namely, that balanced coverage reporting claims from both sides could trigger the perception that the media outlet is receptive to the immoral and threatening side. Future research might investigate this hypothesis by testing whether the

“hostile media” effect occurs only for moralized and threatening media coverage or if it is a more generalized effect.

Last but not least, our findings have implications for the dynamics of political polarization and coalition building. Classical theories of coalition formation (e.g., Axelrod, 1970; Riker, 1962) suggest that moderates should function as pivotal bridge-builders, enabling otherwise distant sides to coordinate. Yet, our findings suggest that moderates may instead be rejected by all sides. Even though structural accounts in political science highlight the potential instrumental value of moderates for coalition formation, our results reveal a psychological barrier such that, instead of being recognized as bridging actors, they are perceived as sympathetic to the opposing side. This divergence between structural and psychological perspectives may help explain why coalition compromise is often elusive in polarized democracies, even when centrist positions exist. Future research might examine whether the stigmatization of moderates not only discourages individual-level moderation but also structurally reduces the feasibility of broad-based coalitions in democratic institutions.

In summary, across five preregistered studies, we demonstrate that individuals with partisan views systematically categorize moderates as out-group members. This “moderate as out-group” effect emerges when issues are moralized and threatening, is amplified among more extreme partisans, and is driven by perceptions that moderates fail to repudiate the opposing side. By identifying the psychological conditions and mechanism underlying this effect, our findings advance understanding of intergroup perceptions and political polarization. They also highlight that positions intended to bridge divides can paradoxically undermine pathways to moderation and consensus.

Constraints on Generality

Our theorizing centers on moralization and perceived threat in sociopolitical contexts. The present studies focused on polarizing sociopolitical issues in the United States. Accordingly, we recruited U.S. participants fluent in English. Although our theory is not U.S.-specific, applications outside the United States will likely require selecting locally moralized and threatening issues and ensuring measurement equivalence. In short, the “moderate as out-group” effect should generalize where the relevant necessary conditions are met, though the issues that meet those conditions may vary by society.

References

- Abramowitz, A. I., & Saunders, K. L. (2006). Exploring the bases of partisanship in the American electorate: Social identity vs. ideology. *Political Research Quarterly*, 59(2), 175–187. <https://doi.org/10.1177/106591290605900201>
- Allport, G. W. (1954). *The nature of prejudice*. Addison Wesley. <https://ia601500.us.archive.org/17/items/in.ernet.dli.2015.188638/2015.188638.The-Nature-Of-Prejudice.pdf>
- Axelrod, R. (1970). *Conflict of interest: A theory of divergent goals with applications to politics*. Markham Publishing Company.
- Bafumi, J., & Shapiro, R. Y. (2009). A new partisan voter. *The Journal of Politics*, 71(1), 1–24. <https://doi.org/10.1017/s0022381608090014>
- Chen, F. S., Minson, J. A., & Tormala, Z. L. (2010). Tell me more: The effects of expressed interest on receptiveness during dialog. *Journal of Experimental Social Psychology*, 46(5), 850–853. <https://doi.org/10.1016/j.jesp.2010.04.012>

- Clifford, S. (2019). How emotional frames moralize and polarize political attitudes. *Political Psychology*, 40(1), 75–91. <https://doi.org/10.1111/pops.12507>
- Durkheim, E. (1973). *Moral education* (E. K. Wilson & H. Schnurer, Trans.). The Free Press. (Original work published 1925)
- Fard, S. H. (2025). Challenging the consensus: The strategic value of homogeneous groups in collective problem solving. *Philosophy of Science*, 93(1), 21–45. <https://doi.org/10.1017/psa.2025.20>
- Feldman, L. (2014). The hostile media effect. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford handbook of political communication* (pp. 549–564). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199793471.013.011>
- Fernbach, P. M., & Van Boven, L. (2022). False polarization: Cognitive mechanisms and potential solutions. *Current Opinion in Psychology*, 43, 1–6. <https://doi.org/10.1016/j.copsyc.2021.06.005>
- Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, 370(6516), 533–536. <https://doi.org/10.1126/science.abe1715>
- Fiorina, M. P., Abrams, S. A., & Pope, J. C. (2008). Polarization in the American public: Misconceptions and misreadings. *The Journal of Politics*, 70(2), 556–560. <https://doi.org/10.1017/s002238160808050x>
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 357–411). McGraw-Hill.
- Fowler, A., Hill, S. J., Lewis, J. B., Tausanovitch, C., Vavreck, L., & Warshaw, C. (2023). Moderates. *American Political Science Review*, 117(2), 643–660. <https://doi.org/10.1017/s0003055422000818>
- Gershon, R., & Fridman, A. (2022). Individuals prefer to harm their own group rather than help an opposing group. *Proceedings of the National Academy of Sciences of the United States of America*, 119(49), Article e2215633119. <https://doi.org/10.1073/pnas.2215633119>
- Giner-Sorolla, R., & Chaiken, S. (1994). The causes of hostile media judgments. *Journal of Experimental Social Psychology*, 30(2), 165–180. <https://doi.org/10.1006/jesp.1994.1008>
- Graham, J., & Haidt, J. (2012). Sacred values and evil adversaries: A moral foundations approach. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 11–31). American Psychological Association. <https://doi.org/10.1037/13091-001>
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–385. <https://doi.org/10.1037/a0021847>
- Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 319–359). Academic Press. [https://doi.org/10.1016/s0065-2601\(08\)60333-0](https://doi.org/10.1016/s0065-2601(08)60333-0)
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295x.108.4.814>
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316(5827), 998–1002. <https://doi.org/10.1126/science.1137651>
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- Haidt, J., & Joseph, C. (2007). The moral mind how five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The Innate Mind* (Vol. 3, pp. 367–391). Oxford University Press. <https://doi.org/10.1093/oso/9780195332827.003.0019>
- Halevy, N., Bornstein, G., & Sagiv, L. (2008). “In-group love” and “out-group hate” as motives for individual participation in intergroup conflict: A new game paradigm. *Psychological Science*, 19(4), 405–411. <https://doi.org/10.1111/j.1467-9280.2008.02100.x>
- Hall, A. B. (2015). What happens when extremists win primaries? *American Political Science Review*, 109(1), 18–42. <https://doi.org/10.1017/s0003055414000641>
- Hamilton, D. L. (1981). *Cognitive processes in stereotyping and intergroup behavior*. Psychology Press. <https://doi.org/10.4324/9781315668758>
- Hartman, R., Hester, N., & Gray, K. (2023). People see political opponents as more stupid than evil. *Personality and Social Psychology Bulletin*, 49(7), 1014–1027. <https://doi.org/10.1177/01461672221089451>
- Haslam, N. (2016). Concept creep: Psychology’s expanding concepts of harm and pathology. *Psychological Inquiry*, 27(1), 1–17. <https://doi.org/10.1080/1047840x.2016.1082418>
- Heltzel, G., & Laurin, K. (2021). Seek and ye shall be fine: Attitudes toward political-perspective seekers. *Psychological Science*, 32(11), 1782–1800. <https://doi.org/10.1177/09567976211011969>
- Hussein, M. A., & Tormala, Z. L. (2021). Undermining your case to enhance your impact: A framework for understanding the effects of acts of receptiveness in persuasion. *Personality and Social Psychology Review*, 25(3), 229–250. <https://doi.org/10.1177/10888683211001269>
- 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*, 153(6), 1425–1448. <https://doi.org/10.1037/xge0001579>
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22(1), 129–146. <https://doi.org/10.1146/annurev-polisci-051117-073034>
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly*, 76(3), 405–431. <https://doi.org/10.1093/poq/nfs038>
- Keltner, D., & Robinson, R. J. (1996). Extremism, power, and the imagined basis of social conflict. *Current Directions in Psychological Science*, 5(4), 101–105. <https://doi.org/10.1111/1467-8721.ep11452765>
- Kim, E., Lelkes, Y., & McCrain, J. (2022). Measuring dynamic media bias. *Proceedings of the National Academy of Sciences of the United States of America*, 119(32), Article e2202197119. <https://doi.org/10.1073/pnas.2202197119>
- Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, 4(3), 279–286. <https://doi.org/10.1038/s41562-019-0766-4>
- Lees, J., & Cikara, M. (2021). Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1822), Article 20200143. <https://doi.org/10.1098/rstb.2020.0143>
- Levendusky, M. S. (2009). The microfoundations of mass polarization. *Political Analysis*, 17(2), 162–176. <https://doi.org/10.1093/pan/mpp003>
- Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, 51(1), 93–120. <https://doi.org/10.1146/annurev.psych.51.1.93>
- Markman, A. B., & Gentner, D. (2001). Thinking. *Annual Review of Psychology*, 52(1), 223–247. <https://doi.org/10.1146/annurev.psych.52.1.223>
- Minson, J. A., & Chen, F. S. (2022). Receptiveness to opposing views: Conceptualization and integrative review. *Personality and Social Psychology Review*, 26(2), 93–111. <https://doi.org/10.1177/10888683211061037>
- Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31(7), 770–780. <https://doi.org/10.1177/0956797620939054>
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review*, 111(3), 781–799. <https://doi.org/10.1037/0033-295x.111.3.781>
- 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. <https://doi.org/10.1177/0146167202286008>

- Rathje, S., Van Bavel, J. J., & Van Der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences of the United States of America*, 118(26), Article e2024292118. <https://doi.org/10.1073/pnas.2024292118>
- Riker, W. (1962). *The theory of political coalitions*. Yale University Press.
- Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naïve realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology*, 68(3), 404–417. <https://doi.org/10.1037/0022-3514.68.3.404>
- Ross, L. (2013). Perspectives on disagreement and dispute resolution. In E. Shafir (Ed.), *The behavioral foundations of public policy* (pp. 108–123). Princeton University Press.
- Ross, L., Lepper, M., & Ward, A. (2010). History of social psychology: Insights, challenges, and contributions to theory and application. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 3–50). Wiley. <https://doi.org/10.1002/9780470561119.socpsy001001>
- Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 27, pp. 255–304). Academic Press. [https://doi.org/10.1016/s0065-2601\(08\)60407-4](https://doi.org/10.1016/s0065-2601(08)60407-4)
- Schwalbe, M. C., Cohen, G. L., & Ross, L. D. (2020). The objectivity illusion and voter polarization in the 2016 presidential election. *Proceedings of the National Academy of Sciences of the United States of America*, 117(35), 21218–21229. <https://doi.org/10.1073/pnas.1912301117>
- Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change*. Yale University Press.
- Siev, J. J., Philipp-Muller, A., Durso, G. R. O., & Wegener, D. T. (2024). Endorsing both sides, pleasing neither: Ambivalent individuals face unexpected social costs in political conflicts. *Journal of Experimental Social Psychology*, 114, Article 104631. <https://doi.org/10.1016/j.jesp.2024.104631>
- Silver, I., & Shaw, A. (2022). When and why "staying out of it" backfires in moral and political disagreements. *Journal of Experimental Psychology: General*, 151(10), 2542–2561. <https://doi.org/10.1037/xge0001201>
- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, 88(6), 895–917. <https://doi.org/10.1037/0022-3514.88.6.895>
- Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Biosocial Science*, 1(S1), 173–191. <https://doi.org/10.1017/s0021932000023336>
- Tappin, B. M., & McKay, R. T. (2019). Moral polarization and out-party hostility in the US political context. *Journal of Social and Political Psychology*, 7(1), 213–245. <https://doi.org/10.5964/jpspp.v7i1.1090>
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324. [https://doi.org/10.1016/s1364-6613\(03\)00135-9](https://doi.org/10.1016/s1364-6613(03)00135-9)
- Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, 49(3), 577–585. <https://doi.org/10.1037/0022-3514.49.3.577>
- Vidal, A. G., & Hunt, C. R. (2026). *The growing influence of social sorting on partisan voting behavior*.
- Wilder, D. A., & Thompson, J. E. (1988). Assimilation and contrast effects in the judgments of groups. *Journal of Personality and Social Psychology*, 54(1), 62–73. <https://doi.org/10.1037/0022-3514.54.1.62>
- Zlatev, J. J. (2019). I may not agree with you, but I trust you: Caring about social issues signals integrity. *Psychological Science*, 30(6), 880–892. <https://doi.org/10.1177/0956797619837948>

Received March 14, 2025

Revision received March 24, 2026

Accepted April 17, 2026 ■