Personality and Social Psychology Bulletin

# Bridging the political divide: Meta-humanization and moral emotions shape partisan reconciliation



| | |
|---|---|
| Journal: | *Personality and Social Psychology Bulletin* |
| Manuscript ID | PSPB-26-156.R1 |
| Manuscript Type: | Empirical Research Paper |
| Keywords: | Meta-humanization, Disgust, Empathy, Anger, Threat, Affective polarization |




| | |
|---|---|
| Abstract: | In an era of deepening political polarization, understanding how partisans believe they are perceived by ideological opponents is critical for explaining both hostility and conciliation. Across five studies (N = 1,541) conducted in the United Kingdom (UK), the United States (US) and Canada, we examine how meta-humanization, perceiving one's political ingroup is seen as fully human by the outgroup, shapes moral emotional responses and downstream intergroup attitudes. Specifically, we test anger, disgust, and empathy as parallel mediators linking meta-humanization to affective polarization, negotiation, teamwork, and willingness for out-party contact, while examining perceived threat to social cohesion as a moderator. Across studies, meta-humanization reliably predicts more conciliatory intergroup attitudes, with disgust emerging as the primary affective pathway. Our findings advance theory on meta-perceptions by demonstrating that meta-humanization shapes political attitudes through moral emotion regulation, and by identifying threat to social cohesion as a condition that, under moderate levels, drives these processes. |

Running title: META-PERCEPTIONS AND PARTISAN RECONCILIATION

# Bridging the political divide: Meta-humanization and moral emotions shape partisan reconciliation

**Word Count: 9969**

**Transparency Statement:** Ethics approval was granted by the author's institution. This research was partially funded through a pump-priming grant from the British Psychological Society. There is no conflict of interest to report. All studies have been registered on the OSF and a blinded peer-review link can be accessed here for the registration: https://osf.io/yvwuc/?view_only=f234ce74f2b54897809db1e172379a11 and for materials and data: https://osf.io/mhqgt/?view_only=941537f4b2364cee87aaebdbf5cba60b

# Bridging the political divide: Meta-humanization and moral emotions shape partisan reconciliation


## Abstract

In an era of deepening political polarization, understanding how partisans believe they are perceived by ideological opponents is critical for explaining both hostility and conciliation. Across five studies ($N$ = 1,541) conducted in the United Kingdom (UK), the United States (US) and Canada, we examine how meta-humanization, perceiving that one's political ingroup is seen as fully human by the outgroup, shapes moral emotional responses and downstream intergroup attitudes. Specifically, we test anger, disgust, and empathy as parallel mediators linking meta-humanization to affective polarization, negotiation, teamwork, and willingness for out-party contact, while examining perceived threat to social cohesion as a moderator. Across studies, meta-humanization reliably predicts more conciliatory intergroup attitudes, with disgust emerging as the primary affective pathway. Our findings advance theory on meta-perceptions by demonstrating that meta-humanization shapes political attitudes through moral emotion regulation, and by identifying threat to social cohesion as a condition that, under moderate levels, drives these processes. Implications for understanding and mitigating identity-based conflicts in polarized political climates are discussed.


**Keywords:** meta-humanization, disgust, empathy, threat, partisan reconciliation

## Transparency Statement

The authors declare no conflict of interest in this project. Materials and data are publicly available on the OSF platform:

Use this link for anonymized OSF registration:

https://osf.io/yvwuc/?view_only=f234ce74f2b54897809db1e172379a11

META-PERCEPTIONS AND PARTISAN RECONCILIATION

Use this link for anonymized materials and data:

https://osf.io/mhqgt/?view_only=941537f4b2364cee87aaebdbf5cba60b

## Introduction

Political polarization has intensified in recent years, particularly in Western democracies such as Canada, the UK, and US (Boxell et al., 2024). This manifests as growing ideological divides between partisans, reduced willingness to compromise, and an erosion of mutual trust (Pew Research, 2022). Scholars have highlighted the role of intergroup dynamics in sustaining and exacerbating these tensions (Renström et al., 2023). Central to these dynamics is the concept of meta-perceptions, or how individuals believe their group is viewed by others. Moore-Berg et al. (2020) demonstrated that meta-perceptions between Democrats and Republicans in the US are strongly negative and incorrectly exaggerated, and that these negative meta-perceptions are amplified by polarization, fostering the perpetuation of intergroup hostility.

In deeply polarized settings, partisanship is not just a matter of differing opinions but often involves moralized dispute, where opposing groups are viewed through moral lenses. Strong moral beliefs tied to partisan issues can intensify polarization, as partisans perceive their opponents not just as ideologically different but as morally deficient (Graham et al., 2009). Partisan identities thus serve as moral frameworks, driving people to view in-party positions as inherently virtuous while casting out-party stances as morally corrupt (Parker & Janoff-Bulman, 2013).

Moral emotions theory provides a foundation for understanding this dynamic (Graham et al., 2009; Rozin et al., 1999). Given the established role of moralized conflicts between partisans

(Garrett & Bankert, 2018), and the relationship between perceptions of immorality and dehumanization (Engels et al., 2024), a critical problem emerges: political partisans may be particularly prone to dehumanizing their opponents driven by meta-dehumanization and moral emotions. Bastian and Haslam (2011) demonstrate that experiences of dehumanization trigger negative emotions (anger and disgust). Whereas, Kteily et al. (2016) demonstrate that meta-dehumanization leads to reciprocal dehumanization and increased support for aggressive policies. This creates a perpetuating cycle of hostility that undermines opportunities for positive intergroup contact, effective teamwork, and broader societal harmony. Thus, moral emotions not only reflect partisan divisions but actively sustain them, making them critical mechanisms in understanding, and potentially mitigating polarization.

From a more optimistic perspective, Pavetich and Stathi (2020) demonstrated that when people perceive their ingroup to be seen as human by the outgroup they show less prejudice and more approach behaviors toward the outgroup. Notably, these effects emerged even under conditions of heightened intergroup threat between Muslims and non-Muslims in Canada and the UK, suggesting that threat does not necessarily impede the benefits of meta-humanization. However, recent work suggests that the relationship between meta-humanization and positive intergroup relations may not be universal. For example, in post-conflict Kosovo, Borinca et al. (2024) found that meta-humanization did not reliably predict willingness to negotiate with the outgroup.

Taken together, these findings suggest that the impact of meta-humanization may depend on situational factors, for example salience of threat, that shape its psychological significance. Perceived threat should differ substantially across active conflict, post-conflict, and relatively stable intergroup contexts. Pavetich and Stathi (2020) observed that meta-humanization effects

were most consistent when high-threat was salient (e.g., during heightened concern about terrorism), however in a context with diminished salience of realistic threat (media focus had shifted to other sociopolitical events), the findings became inconsistent. This pattern implies that threat may not dampen meta-humanization's effects, but rather condition when perceptions of being seen as human become especially meaningful for intergroup attitudes.

In the current research we investigate threat to social cohesion, referring to the threat of political divisions eroding shared norms, mutual respect, and a sense of collective belonging, essential for cooperative functioning in society (see Packer & Ungson, 2024). We propose that this particular type of threat may shape the interplay between meta-humanization, moral emotions, and out-party attitudes. Across four studies (Studies 2a-3b), we examined whether threat to social cohesion moderates the relationship between meta-humanization and moral emotions (anger, disgust, and empathy), and how these processes relate to affective polarization and conciliatory attitudes, *see Figure 1.* By integrating research on meta-perceptions, moral emotions, and social cohesion (Giner-Sorolla & Russel, 2019; Packer & Unger, 2024; Pavetich & Stathi, 2020; Scatolon et al., 2023), we offer a framework for understanding when and why meta-humanization fosters cross-party reconciliation, highlighting the moderating role of threat and the mediating role of moral emotions.

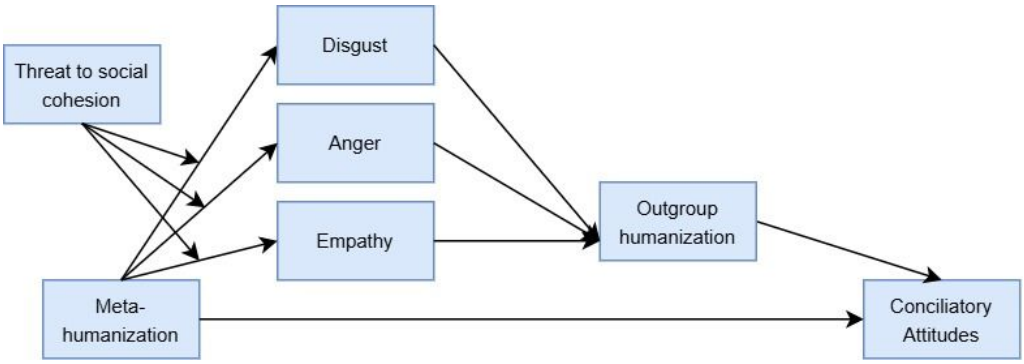


*Figure 1.* Proposed conceptual model of the effect of meta-humanization on conciliatory attitudes, mediated through moral emotions (disgust, anger, and empathy) and outgroup humanization, with the path from meta-humanization to moral emotions moderated by perceived threat to social cohesion, while controlling for out-partisan contact



**Meta-(de)Humanization and Moral Emotions**

Moral emotions, including anger, disgust, and empathy, play a central role in shaping intergroup perceptions, including the extent to which groups humanize or dehumanize one another (Giner-Sorolla & Russell, 2019; Scatolon, 2023). Anger typically arises in response to perceived injustice, motivating confrontation and punishment, whereas disgust is linked to violations of purity norms and often drives avoidance and dehumanization (Tybur et al., 2020; Landry et al., 2022). Empathy, in contrast, fosters prosocial engagement and reconciliation by promoting perspective-taking and concern for others (Dovidio et al., 2010; Vanman, 2016). Although prior research has independently linked (de)humanization to moral emotions (Giner-Sorolla et al., 2023) these literatures have not been integrated as in the current research.

According to the prophylactic model of dehumanization disgust is a potent predictor of dehumanization processes (Landry et al., 2022). Landry and colleagues suggest that the behavioral immune system motivates people to avoid perceived vectors of disease (outgroups, i.e. immigrants, Chinese people during the COVID-19 pandemic, and obese people to name a few) by eliciting disgust, which in turn promotes dehumanizing responses. Disgust thus serves as an inhibitor of perceived humanity, and behaviorally motivates social distance and exclusion (Tybur et al., 2020). This work implies that if meta-humanization was to improve intergroup relations, it may do so largely by reducing contamination-based disgust.

By contrast, the role of anger in intergroup settings is more nuanced. Anger typically arises when individuals perceive intentional mistreatment or injustice (Tybur et al. 2020; Giner-Sorolla & Russell, 2019). In the context of meta-perceptions, anger can motivate retaliatory intentions (Owuamalam et al., 2013), while sometimes anger can motivate prosocial actions to restore justice (van Doorn et al., 2014), it is more sensitive to perceptions of deliberate harm rather than moral contamination, and less directly tied to the purity-based pathways through which meta-humanization may exert its effects.

Moral emotions can provide pathways to conciliation if framed around shared values. Empathy, the capacity to understand and share the feelings of others, is a well-established contributor to positive intergroup attitudes (Dovidio et al., 2010). When people feel their group is humanized by the outgroup, they are more likely to extend empathy toward the outgroup, fostering recognition of shared humanity (Prati et al., 2023). This emotional connection can mitigate the impact of prior dehumanization, facilitating conciliatory attitudes in polarized contexts (Moore-Berg et al., 2021). Accordingly, while empathy remains theoretically relevant

(Scatolon, et al., 2023), prior work has not positioned it as the mechanism that meta-humanization exerts its effects.

In this research, we test the effect of meta-humanization on disgust, anger and empathy as drivers of the pathway to cross-party conciliatory attitudes. We argue that understanding the impact of meta-humanization on these moral emotions is critical in explaining how meta-humanization shapes its outcomes. If enhanced empathy, and reduced anger and disgust are in fact emotional responses that meta-humanization elicits, then this should foster feelings of both outgroup humanization as well as attitudes centered around willingness for positive intergroup behaviours such as teamwork, cooperation, and negotiation.

**Threat to Social Cohesion**

Threats to social cohesion constitute a socially rooted form of threat that activates moral and emotional defenses by undermining shared values, trust, and collective belonging (Onraet et al., 2014; Vonhm, 2025). Threats of this kind commonly stem from intergroup conflict, ideological polarization, and socioeconomic disparities, all of which can fracture social cohesion (Renström et al., 2023). In political contexts, threats of this nature often manifest through opposing political factions. For example, polarization between political parties can erode interpersonal trust and cooperative norms, exacerbating societal fragmentation and threat (Lee, 2022).

The relationship between threat and intergroup behavior is not uniformly negative. Under certain circumstances, perceived threats can paradoxically motivate constructive engagement with outgroups (Jonas et al., 2014). This dual effect has been observed in various conflict contexts, where heightened awareness of shared risks or challenges prompts individuals to seek

dialogue or collaboration (Berendt et al., 2023), as well as evidence that meta-humanization improved desire for contact under perceived high threat (between Muslims and non-Muslims during a wave of terrorist attacks).

Building on this, the present research is among the first to explicitly integrate meta-humanization with moral emotion theory, identifying associations between moral emotions as a key psychological mechanism linking meta-perceptions to partisan reconciliation under threat to social cohesion. We propose that threat acts as a moderator, amplifying the effect of meta-humanization on moral emotions in high-threat political contexts. When social cohesion feels endangered, such as during intense partisan conflicts, individuals may become more attuned to the humanity of the ingroup and outgroup, enhancing the emotional impact of meta-humanization. We suggest that meta-humanization will interrupt this process, when perceived threat is high. By affirming that the outgroup sees one's group as fundamentally human and morally legitimate (i.e. the process of meta-humanization), pathways for cooperative engagement should be enhanced. This moderated relationship could illuminate the conditions under which meta-humanization interventions most effectively bolster social cohesion, offering theoretical and practical insights for mitigating political division.



## The Current Research

Proposing an innovative framework that takes into consideration meta-perceptions, moral emotions and threat literature, we seek to extend our understanding of when and how meta-humanization exerts its effect in contexts of political polarization. Despite growing evidence that meta-humanization can reduce intergroup hostility (e.g., Borinca et al., 2024; Pavetich & Stathi,

2020), a deeper understanding of how these effects unfold at the emotional level is needed. In particular, the affective mechanisms through which feeling humanized by the outgroup shape intergroup attitudes have received limited empirical attention, with moral emotions remaining largely untested within this framework. The present research addresses this gap by positioning moral emotions as central pathways through which meta-humanization may foster conciliatory orientations. Further, we single out the role of perceived threat to social cohesion as a particularly relevant moderator of the pathway from meta-humanization to cross-party conciliatory attitudes. By clarifying the affective processes through which meta-humanization exerts its effects, and testing whether these processes operate similarly across variation in perceived threat we provide a more comprehensive account of the psychological dynamics that shape partisan attitudes in polarized contexts.

We test our hypotheses in three Western democracies (Canada, UK, US), seeking to provide a theoretical model that explains when, and how, meta-humanization leads to moral emotions motivating attitudes that align with ethical norms and ameliorate social divisions. By contextualizing this research in political partisanship, the current work extends existing research to a critical and timely context. While these countries share common democratic traditions, they differ in their political systems, cultural norms, and degrees of polarization. Comparing these contexts allows for a more nuanced understanding of how meta-humanization and threat perception operate across different political environments. Moreover, our research explores the affective mechanisms—specifically empathy, anger, and disgust—through which meta-humanization shapes attitudes, providing new insights into the psychological processes underlying cross-partisan reconciliation. Finally, these studies investigate the role of perceived

threat to social cohesion, offering a deeper understanding of how these perceptions interact with moral emotions to both hinder and promote constructive cross-party engagement.

Taking into consideration previous research demonstrating that intergroup contact can reduce prejudice and improve humanity attributions (Paolini et al., 2021) we controlled for existing out-partisan contact, aiming to demonstrate the unique effects of meta-humanization, above and beyond existing contact. In Study 1 we extend previous research (e.g., Giner-Sorolla et al., 2019; Pavetich & Stathi, 2020; Scatolon, 2023) hypothesizing that meta-humanization activates moral emotions (anger and disgust), reduces hostile behavioural intentions, and decreases social distance (H1). We conducted this research in the UK, in the context of Conservative and Labour party partisans in February 2021. At the time the Covid-19 pandemic largely dominated the news, and no major political events occurred.

In Studies 2a and 2b we expected that participants exposed to meta-humanization would report reduced affective polarization and greater conciliatory orientations toward the out-party, reflected in higher support for negotiation, teamwork, and intergroup contact, relative to control and meta-dehumanization conditions (H2). Thirdly, we investigated a conditional process, hypothesizing that when intergroup threat is perceived to be high, it will moderate an indirect effect of meta-humanization, via moral emotions (in parallel) and outgroup humanization (H3). In line with previous literature on humanization and prejudice (Capozza et al., 2017), and literature on meta-humanization and threat (Pavetich & Stathi, 2020), we expected perceived threat to act as a contextual amplifier condition, such that the positive effects of meta-humanization on moral emotions and intergroup outcomes would be amplified under high levels of threat. This research is conducted in the more aggressively polarized context of US Democrat and Republican partisans one week prior to the 54$^{th}$ Presidential elections.

In Studies 3a and 3b we replicate Studies 2a and 2b in a different sociopolitical context, between Canadian Liberal and Conservative partisans. We predict that the findings should mirror those in Study 2a and 2b, given the contentious political context affects Canada as well. Studies 2a-3b were pre-registered on the OSF platform osf.io/68yht.

## Study 1

### Method

**Participants and Design.** Data were collected from Prolific, participants were paid the equivalent of £9.00/hr for the 8-min study. Participants were required to consider themselves as British, living in the UK, and identify as supporters of the Labour or Conservative parties. We captured a representative sample of 353 participants (*M* age = 35.62, *SD* = 15.62); 61.1% female; predominantly White ethnicity 79.9%, identifying with either Conservative 52.0% or Labour 48.0% ideology. An a-priori power analysis was conducted using G*Power 3.1 to detect the predicted meta-perception × political orientation interaction, assuming a small effect ($f = .15$), we required $N = 351$ participants with an $\alpha = .05$ and power of .80. Due to constraints, we collected a sample size of 327. A sensitivity power analysis further indicated that with a total sample size of $N = 327$, $\alpha = .05$, and power = .80, the present study was sufficiently powered to detect effects of $f = .16$ or larger.

**Procedure.** Participants were asked to complete a questionnaire as part of research on social attitudes, beginning with demographic questions. Embedded within the demographics was a measure asking participants how frequently they have contact with the political outgroup, ranging from (1) *extremely infrequently* to (5) *extremely frequently.* We used this measure to control for out-partisan contact across all studies.

After filling out the demographic information, participants were randomly assigned to one of the two levels in the experimental condition: (meta-dehumanization vs. meta-humanization). Based on whether participants self-reported as Labour of Conservative the meta-perception manipulation was tailored to suit the perspective of the outgroup. The manipulation was presented as an excerpt from a series of Twitter (now X) posts. The excerpt discussed the dehumanizing or humanizing perceptions of opposing political partisans. For example, in the meta-dehumanization condition participants read a series of mock Twitter posts that described either Labour or Conservative partisans as "animals and brutes lacking in self-control." In the meta-humanization condition participants read the same randomized series of mock Twitter posts however the posts described Labour or Conservative partisans as "sophisticated."

Immediately following the experimental manipulation, participants were asked manipulation check questions that tapped on the unique aspects of the respective condition that they were exposed to. Participants then completed the counterbalanced dependent measures: dehumanization, hostile behavioural intentions, social distance, anger, and disgust.

## Materials

### *Meta-perception manipulation check.*

To ensure the effectiveness of the manipulation, we assessed the extent to which participants perceived that they were dehumanized or humanized in the Twitter posts they read. Participants were asked to indicate their level of agreement for two items ($r$ = .56), for example, "Labour party supporters were referred to as animals in the Twitter posts I just read," in the meta-dehumanization condition.

***Dependent measures.*** All measures, unless otherwise stated, were measured on a seven-point Likert scale ranging from (1) *strongly disagree* to (7) *strongly agree*, where higher scores indicated higher agreement with the items.

*Hostile behavioral intentions*. The intention to cause harm or intentional suffering toward partisans was examined using a measure adapted from Golec de Zavala et al. (2013). Participants were asked, "To what extent did reading the Twitter posts make you want to: hurt, offend, injure, intimidate, and humiliate Conservatives" ($\alpha$ = .93). Each of the five words were presented and answered individually.

*Social distance*. To assess desire for distance from the outgroup a three-item measure adapted from Moore-Berg et al. (2020) was used. An example item is, "How comfortable would you feel if your doctor was (Labour/Conservative)?" ($\alpha$ = .94).

*Moral emotion ratings*. To measure participants' emotions, they were asked to rate how much anger and disgust they experienced, after reading the Twitter posts, adapted from Tapias et al. (2007). Items were presented individually.

**Results**

**Manipulation checks**. We began by examining whether the meta-perception manipulation influenced how participants believed they were viewed by the outgroup (Labour or Conservative partisans). Those exposed to the meta-humanization were significantly more likely to perceive that they were humanized by the outgroup ($M$ = 4.25, $SD$ = 1.06), than those in the meta-dehumanization condition ($M$ = 1.51, $SD$ = 1.02), $t$(352) 24.78, $p$ < .001.

**Main analyses**. A 2 (partisan identity: Labour, Conservative) x 2 (meta-perception: meta-dehumanization, meta-humanization) ANCOVA was performed on the dependent

measures, controlling for out-partisan contact. Significant interactions between partisan identity and meta-perceptions were only observed for anger, however main effects of the meta-perception manipulation were observed across all dependent variables, see Table 1. The interactions were dissected using simple main effects to keep the error term constant. Bonferroni corrected alpha values were adjusted to .025 to reduce familywise Type I error for multiple comparisons in the post hoc analysis. The results indicated a statistically significant simple effect for Conservative supporters exposed to the meta-perception manipulation for anger $F(1, 349) = 52.27$, $p < .001$, $\eta^2_p = .13$. See Table 2 for a summary of post hoc tests.

As expected, there were significant differences between meta-dehumanization and meta-humanization across all dependent measures for participants in both the Labour and Conservative conditions. Contrary to predictions, only an interaction between political affiliation and meta-perception exposure was observed for anger. Although there were no significant differences for hostility, social distance, and disgust between partisans, the findings are in line with previous literature suggesting partisans share similar out-party attitudes (Moore-Berg et al., 2020). These findings are informative in guiding Studies 2a-3b as they demonstrate more importantly that meta-(de)humanization causes an effect in moral emotions directed to the out-party.



## Study 2a

The results of Study 1 informed the rationale for the subsequent studies in three ways. Firstly, after acquiring initial evidence for the effectiveness of meta-(de)humanization on moral emotions, we tightened up the methodology, using a control group as a point of comparison to the meta-perceptions. Secondly, we aimed to test whether the results were consistent with previous theorizing related to threat potentially being a catalyst to conciliatory attitudes (Borinca et al., 2024), therefore we included threat as a moderator, see Figure 1. Finally, we chose to

include empathy alongside anger and disgust as parallel mediators in the model as previously mentioned, empathy is related to conciliatory attitudes thus we deemed it particularly pertinent (Moore-Berg et al., 2021).

## Method

**Participants and design.** Data were collected again from Prolific. Participants were required to consider themselves as American, living in the US, and identify with being supporters of the Republican party. The sample consisted of 301 participants (*M* age = 47.02, *SD* = 13.58); 63.8% female; predominantly White ethnicity 83.1%. Participants reported average partisanship identification *M* = 3.45, *SD* = 1.08. An a priori Monte Carlo power analysis (5,000 simulations, seed = 123) was conducted for the moderated serial mediation model, with perceived threat moderating the IV → emotions path. Assuming conservative small effect sizes for all paths (a1 = .20, a2 = .15, d21 = .20, b1 = .20, b2 = .20), a sample size of N = 300 would provide an estimated power of .97 at the mean level ($\alpha$ = .05), and .99 at high levels of the moderator, consistent with our theoretical predictions. Analyses were conducted using a custom R simulation script, available in the supplementary materials. Sensitivity power analyses were conducted using G*Power to determine the smallest effects detectable given the final sample sizes ($\alpha$ = .05, $\beta$ = .80). Across studies, we were sufficiently powered to detect small effects for the direct and mediational paths linking meta-humanization to moral emotions and downstream outcomes ($f^2 \approx .02$), as well as small interaction effects for the moderation of these paths by perceived threat ($\Delta R^2 \approx .02$). These analyses indicate that the present studies were adequately powered to detect effects of the magnitude typically observed in intergroup and political psychology research.

This study was designed using SurveyLab and was conducted online in one session. Participants were randomly assigned to one of three conditions, that is, meta-dehumanization versus meta-humanization versus control. In the experimental conditions, participants read the same style fictional X posts as in Study 1, however they were phrased to demonstrate either humanizing or dehumanizing attitudes toward Republicans purportedly from Democrats. Participants in the control condition did not view any X posts, they proceeded to answer the dependent measures.

**Procedure.** After collecting demographic information, participants completed a measure of threat to social cohesion. Participants were then told that, "We recently conducted a survey about online social attitudes, similar to the one you are completing now. In that survey, we collected posts from X and looked at how Democrats felt toward Republicans. Please read the X posts, and then we will have some follow-up questions for you." After reading the randomly assigned posts, participants were asked manipulation check questions. All participants then completed measures of moral emotions and humanization, followed by randomized presentation of affective polarization, negotiation, compromise, teamwork, and willingness for contact.

**Measures.** All measures, unless otherwise stated, were measured on a seven-point Likert scale ranging from (1) *strongly disagree* to (7) *strongly agree*, where higher scores indicated higher agreement with the items.

***Frequency of outgroup contact.*** Frequency of contact was assessed with the same one-item measure as in Study 1.

***Threat to social cohesion.*** The perception participants felt that there was a threat to the shared values in their community between Democrats and Republicans was measured by three-

items adapted from Onraet et al. (2014), for example "The values in our society have gone seriously off track," ($\alpha$ = .86).

***Meta-perception manipulation check.*** The manipulation check was assessed with the same three items as in Study 1 ($\alpha$ = .95).

***Humanization.*** To measure participants' humanization toward Republicans the Ascent of Man scale was used (Kteily et al., 2016). Responses were measured on a scale ranging from 0 (highly uncivilized) to 10 (highly civilized). Participants were asked to use a slider to indicate how evolved they believed Democrats to be.

***Moral emotions.*** The moral emotions of anger, disgust, and empathy were assed with the same one-item measures respectively as in Study 1.

***Affective polarization.*** To capture political animus toward the outgroup, 11-items adapted from Druckman et al. (2020) were used, for example, "Democrats are open-minded". Items 1-5 and 9-11 were reverse scored so that higher scores indicated stronger feelings of politically charged prejudice toward the outgroup ($\alpha$ = .94).

***Openness for contact.*** Three items were used to measure openness for future out-partisan contact. These were derived from Borinca et al. (2024), for example, "In general, are you willing to have contact with Democrats in the future?" ($\alpha$ = .88).

***Teamwork.*** Three-items were adapted from Varela & Mead (2018) to measure individuals' desire to work well with the opposing political party in order to achieve a common goal, for example, "In a team, Democrats and Republicans should emphasize common goals rather than dwelling on differences" ($\alpha$ = .91).

***Negotiation.*** We assessed participants' support for their political party to negotiate with the adversarial party to reach a solution using two items adapted from Kalisi (2021) "Do you think that Republicans should make a concerted effort to negotiate resolutions with Democrats" ($\alpha$ = .93).

***Compromise.*** Intergroup compromise within a political setting was measured with eight-items adapted from Willis et al. (2017) demonstrating the desire to achieve results that are acceptable to both political parties. An example item is, "If the other party's solution is proven to be better for the country, we should help them" ($\alpha$ = .95).

**Results**

**Manipulation Checks.** An ANOVA was performed on meta-humanization revealing that the main effect of our experimental manipulation was significant, $F(2, 293) = 207.48$, $p < .001$, $\eta^2_p = 0.59$. LSD comparisons showed that participants perceived that the outgroup humanizes their ingroup more in the meta-humanization condition ($M = 5.31$, $SD = 1.52$) than in the meta-dehumanization condition ($M = 1.52$, $SD = 0.88$), $p < .001$ and control condition ($M = 3.91$, $SD = 1.23$), $p < .001$. There was also a significant difference between meta-dehumanization and the control condition, $p < .001$.

**Conditional Process Analysis**. We tested the effects of the experimental manipulation (meta-humanization vs. meta-dehumanization vs. control) on the dependent variables (affective polarization, support of intergroup negotiation, compromise, teamwork, and intergroup contact) via the mediators (empathy, anger, and disgust – in parallel), and whether threat moderates the effects. Those analyses were conducted using Model 85 in the PROCESS Macro for SPSS

(Hayes, 2018) with 5000 bootstrapped samples and 95% bias-corrected confidence intervals for the indirect effects at 16th, 50th, and 84th percentiles of the moderator (threat).

PROCESS indicator coding was used to construct products for a comparison of the effect of meta-humanization to meta-dehumanization and the control group in the hypothesized model. First, using meta-dehumanization as the reference point to meta-humanization (Model 1), and second, using the control group as the reference point to meta-humanization (Model 2) as this compares the mean difference between the two focal predictors in each model. The direct and indirect effects revealed by our analysis are reported in unstandardized coefficients controlling for out-partisan contact in Table 3.

In line with predictions there was a significant meta-perception by threat interaction in both Model 1 and Model 2. An inspection of the conditional effects of the focal predictor at values of the moderator indicated that among those with moderate (average) and high levels of threat, meta-humanization significantly reduced affective polarization and improved conciliatory attitudes through reduced disgust and increased outgroup humanization compared to those in the meta-dehumanization and control condition. Contrary to predictions, the conditional process of meta-humanization was not significant through reduced anger or empathy. For brevity, only significant results will be reported, see Table 4.

In summary, we observed that participants in the meta-humanization condition, compared to those in other conditions, demonstrated reduced disgust, increased humanization, reported reduced affective polarization and increased support for negotiation, teamwork, cooperation, and greater openness to out-partisan contact, under moderate levels of threat. However, this pattern did not hold true for the other anger and empathy. To replicate these findings, accounting for the political opposition's perspective to test our model, we conducted Study 2b.

To assess the robustness of our findings across all four studies, we tested two alternative models including reverse association and alternative predictor orderings; none provided comparable support for our hypothesized mediation pathway. Additionally we have included main and interaction effects analyses for transparency across studies (see Supplementary Materials for full results).

## Study 2b

In Study 2b we replicated the experimental design of Study 2a by considering the same partisan context but with a different sample. Specifically, we used American Democrats as the participant sample under investigation to provide a comparative approach and demonstrate that the phenomena under investigation share similarities.

### Method

**Participants and design.** Data were collected from Prolific and participants were required to consider themselves as American, living in the US, and identify with being supporters of the Democrat party. The sample consisted of 302 participants ($M_{age}$ = 46.38, *SD* = 14.02); 66.6% female with predominantly White ethnicity 79.9%. Participants' democratic partisanship was reported as being slightly above average (*M* = 3.76, *SD* = 1.11).

**Outcome measures.** All measures were identical to those used in Study 2a with alpha reliabilities of: Threat to social cohesion ($\alpha$ = .81), affective polarization ($\alpha$ = .92), negotiation ($\alpha$ = .86), compromise ($\alpha$ = .90), desire for contact ($\alpha$ = .85), and teamwork ($\alpha$ = .89).

### Results

**Manipulation Checks.** An ANOVA was performed on meta-humanization revealing that the main effect of our experimental manipulation was significant, $F(2, 299) = 257.76$, $p < .001$, $\eta^2_p = 0.63$. LSD comparisons showed that participants perceived that the outgroup humanizes their ingroup more in the meta-humanization condition ($M = 5.73$, $SD = 1.10$) than in the meta-dehumanization condition ($M = 1.70$, $SD = 1.08$), $p < .001$, 95 % CI [3.68, 4.38] and control condition ($M = 3.43$, $SD = 1.55$), $p < .001$, 95 % CI [1.94, 2.64]. There was also a significant difference between meta-dehumanization and the control condition, $p < .001$, 95 % CI [− 2.09, – 1.39].

**Conditional Process Analysis.** Analyses were conducted identically to in Study 2a. We observed a significant meta-perception by threat interaction in both Model 1 and Model 2, again only through disgust. An inspection of the conditional effects of the focal predictor at values of the moderator indicated that among those with moderate and high levels of threat, meta-humanization significantly reduced disgust and reciprocated humanization compared to those in the control condition (Model 2). Whereas compared to meta-dehumanization (Model 1) this effect was only significant under moderate levels of threat through disgust and outgroup humanization in reducing affective polarization, and improving negotiation, cooperation, and desire for contact, not teamwork, see Table 5. Similar to Study 2a the index of moderated mediation for the indirect effect of meta-humanization through outgroup humanization was only statistically significant through disgust, not anger or empathy, indicating that threat moderated the indirect effect of meta-humanization through disgust and outgroup humanization on reduced affective polarization and improved conciliatory attitudes, see Table 6.

In summary, Study 2b replicated Study 2a in the same political context using a different ingroup, American Democrats, in order to provide perspectives from opposing political partisans.

Similar effects were seen across Study 2a - 2b albeit weaker across all outcome variables in both Model 1 and Model 2. The findings from Study 2b replicated those in Study 2a in two ways, firstly to demonstrate that under moderate levels of threat meta-humanization improved conciliatory attitudes and reduced affective polarization through reduced disgust and improved outgroup humanization. Secondly, that these effects were not seen in parallel for anger and empathy.

## Study 3a

To test generalizability of the results obtained in Studies 2a and 2b under different contexts and support our hypothesized model, Study 3a replicated the experimental design in a partisan context but with a different sample. Specifically, we used Canadian Conservatives to provide a comparative approach and demonstrate the validity of our model. Although the Canadian context may provide a somewhat less contentious political context, compared to Study 2, that was implemented a few days before the 2024 United States Federal Election, research has demonstrated rising tensions between partisans in Canada (Merkley, 2023). In line with this, we predicted that the results from Study 3a will replicate those in Study 2.

### Method

**Participants and design.** Data collection processes and inclusion criteria were identical to Study 2, with the exception that participants were required to consider themselves politically as Conservatives, living in Canada. The sample consisted of 310 participants ($M_{age}$ = 46.22, *SD* = 16.70; 55.4% male. The ethnic background of participants was predominantly 73.1% White. The study was administered using Qualtrics Panel paid platform.

The methodology of Study 3a was identical to that of Study 2. Participants answered the same demographics, questionnaires, and were randomly exposed to the same study conditions; however, all study material was phrased to suit Conservatives. Alpha reliabilities for the measures were: Threat ($\alpha$ = .89), manipulation check ($\alpha$ = .90), compromise ($\alpha$ =.91), negotiation ($r$ = .65, $p$ = .001), intergroup contact ($\alpha$ = .84), teamwork ($\alpha$ = .90), and affective polarization ($\alpha$ = .90).

**Results**

**Manipulation checks.** An ANOVA was performed on meta-humanization revealing that the main effect of our experimental manipulation was significant, $F(2, 265) = 28.15$, $p < .001$, $\eta^2_p = 0.19$. LSD comparisons showed that participants perceived that the outgroup humanizes their ingroup more in the meta-humanization condition ($M$ = 4.55, $SD$ = 1.35) than in the meta-dehumanization condition ($M$ = 2.92, $SD$ = 1.63), $p < .001$, 95 % CI [1.21, 2.07] and control condition ($M$ = 3.81, $SD$ = 1.23), $p < .001$, 95 % CI [.32, 1.16]. There was also a significant difference between meta-dehumanization and the control condition, $p < .001$, 95 % CI [− 1.31, – .48].

**Conditional Process Analysis.** We used the same analytical method as in Study 2a. As predicted, there was a significant meta-perception by threat interaction in both Model 1 and Model 2, through disgust on all outcome variables. The conditional effects of the focal predictor at values of the moderator indicated that among those with moderate and high levels of threat, meta-humanization significantly reduced disgust and reciprocated humanization compared to those in the meta-dehumanization and control condition, see Table 7. The index of moderated mediation for the indirect effect of meta-humanization through outgroup humanization again was significant through disgust, indicating that threat moderated the indirect effect of meta-

humanization through disgust and outgroup humanization on reduced affective polarization and improved conciliatory attitudes, see Table 8.

These findings are consistent with Study 2 indicating that threat moderated the indirect effect of meta-humanization through disgust on outgroup humanization, improving conciliatory attitudes compared to meta-dehumanization and the control group. Similar to the previous studies, we found the effect being moderated consistently through moderate levels of threat.

## Study 3b

To provide a cohesive replication of our work we conducted the final study in this series in Canada, capturing the perspective of those who identify as supporting Liberal political ideology with the out-party being Conservatives.

### Method.

**Participants and design.** Data collection processes and inclusion criteria were identical to Study 3a, with the exception that participants were required to identify as Liberal, living in Canada. The sample consisted of 300 participants ($M_{age}$ = 48.61, $SD$ = 16.32; 51.5% male). The ethnic background of participants was predominantly 71.1% White.

The methodology of Study 3b was identical to that of Study 3a; however, all study material was phrased to suit Liberals from Canada rather than Conservatives. Alpha reliabilities for the measures were: Threat ($\alpha$ = .86), manipulation check ($\alpha$ = .88), compromise ($\alpha$ =.92), negotiation ($r$ = .80, $p$ = .001), intergroup contact ($\alpha$ = .87), teamwork ($\alpha$ = .91), and affective polarization ($\alpha$ = .90).

### Results

**Manipulation checks.** An ANOVA was performed on meta-humanization revealing that the main effect of our experimental manipulation was significant, $F(2, 265) = 44.82$, $p < .001$, $\eta^2_p = 0.25$. LSD comparisons showed that participants perceived that the outgroup humanizes their ingroup more in the meta-humanization condition ($M = 4.64$, $SD = 1.57$) than in the meta-dehumanization condition ($M = 2.64$, $SD = 1.46$), $p < .001$, 95 % CI [1.58, 2.41] and control condition ($M = 3.91$, $SD = 1.23$), p < .001, 95 % CI [.31, 1.15]. There was also a significant difference between meta-dehumanization and the control condition, $p < .001$, 95 % CI [− 1.69, – .85].

**Conditional Process Analysis.** We used the same analytical method as in Study 2a. A significant meta-perception by threat interaction was observed in both Model 1 and Model 2. We found the effect consistently through disgust on all outcome variables in Model 1 and Model 2, and through empathy only in Model 2, see Table 9. An inspection of the conditional effects of the focal predictor at values of the moderator indicated that among those with low or moderate levels of threat, meta-humanization significantly reduced disgust and reciprocated humanization compared to those in the meta-dehumanization and control condition. The index of moderated mediation for the indirect effect of meta-humanization through outgroup humanization was significant through disgust (Model 1 and 2) and empathy (Model 2) in parallel, indicating that threat moderated the indirect effect of meta-humanization through outgroup humanization on reduced affective polarization and improved conciliatory attitudes, see Table 10.

In summary, the findings from Study 3b replicated support for our proposed model to indicate that threat moderated the indirect effect of meta-humanization through reduced disgust and increased outgroup humanization, improving conciliatory attitudes compared to meta-dehumanization and the control group. Interestingly, in this final study we also observed this

effect through improved empathy, however only in Model 2 (meta-humanization compared to the control condition).

## Discussion

Across five studies spanning the UK, Canada, and the US, we advance the understanding of political partisanship by identifying disgust as an affective mechanism through which meta-humanization shapes conciliatory attitudes and affective polarization. Our findings demonstrate that perceptions of being humanized by political opponents systematically regulate the emotional response of disgust, and not empathy or anger, which in turn predicts affective polarization, willingness to negotiate, teamwork, and desire for intergroup contact. In doing so, this work extends prior research on threat and meta-humanization by moving beyond whether meta-perceptions matter to clarify how they operate emotionally, and when their effects are most likely to emerge. By integrating meta-perception theory with moral emotion frameworks, this research offers a account of how partisan divides may be softened even in highly polarized Western democracies, where political conflict increasingly threatens social cohesion (Boxell et al., 2024; Pew Research Center, 2022).

Study 1 documented that morally entrenched emotions (anger and disgust) are an outcome of meta-humanization and meta-dehumanization, the same types of emotions that have been shown to be associated with dehumanization (Giner-Sorola & Russell, 2019). This initial study guided us in reframing our focus in the subsequent studies in this project. Empathy was not included in Study 1, however through further research while conceptualizing the subsequent studies (2a-3b), empathy stood out to us as being particularly relevant to the context and outcomes we were investigating (Moore-Berg et al., 2021).

Across studies 2a-3b, we extend the findings from Study 1 by testing a conditional process model in which threat interacts with meta-humanization to shape moral emotions, after controlling for out-partisan contact. Specifically, we examined anger, disgust, and empathy in parallel as pathways to outgroup humanization, and in turn, conciliatory attitudes. Consistently across studies the findings supported our prediction that meta-humanization reduced disgust, and that disgust mediated the relationship between meta-humanization and conciliatory attitudes, under conditions of threat. Disgust emerged as a consistent mediator across studies, whereas anger was largely null and only in Study 3b was empathy a significant emotional driver of meta-humanization when compared to the control group (Model 2). Importantly, we argue that this pattern should not be interpreted as a failure of the broader moral-emotional framework, but rather as evidence of emotional specificity in the operation of meta-humanization. Disgust is implicated in processes of moral exclusion, contamination, and dehumanization (Giner-Sorolla et al., 2023), making it especially sensitive to perceptions of being humanized by an outgroup. In contrast, anger is typically driven by appraisals of blame, intentional harm, and injustice—appraisals that may not be directly altered by meta-humanization alone (Tybur et al., 2020). Empathy, while theoretically relevant, is a more cognitively and motivationally demanding emotion that may only emerge under particular conditions, for example, involving intergroup contact (Prati et al., 2023) or among empathy-inclined individuals (Scatolon et al., 2023). Taken together, these findings suggest that meta-humanization primarily attenuates moral boundary emotions (such as disgust) rather than uniformly amplifying prosocial affect.

It is worthwhile to note that some variation in the moderator (threat to social cohesion) was observed among Studies 2a-3b. In Study 2b (American Democrat) and 3b (Canadian Liberals) the interaction of threat and meta-humanization was significant between low to

moderate levels of threat, whereas in Study 2a (American Republican) and 3a (Canadian Democrats) the interaction was significant at moderate to high levels of threat. Despite these variations, we argue that this finding is theoretically informative. Onraet et al. (2014) review literature demonstrating right-wing attitudes to be a predisposition leading to heightened sensitivity to perceive threat (e.g., Cohrs, 2013; Feldman & Stenner, 1997; Stephan & Renfro, 2002). Moreover, there is evidence, for example, that contact works best for people who are more prejudiced or report higher social dominance orientation and authoritarianism (see Turner et al., 2020). It could be that those with more conservative ideology may also be more inclined to benefit from the process of meta-humanization, which show the 'good will' of the outgroup.

The variation in moderation could also suggest that the meaning of threat might differ slightly across samples for two reasons. Firstly, threat may act as a protective buffer. When threat is perceived to be low, people might feel safe and open enough that they can benefit from meta-humanization; however, at high levels of threat, people may be too defensive, fearful, or closed off, meaning meta-humanization has less effect because people are already in a defensive posture. This is consistent with threat defense theories, which posit that high threat can inhibit openness (Jonas et al., 2014). In contrast, we suggest that meta-humanization may exert positive intergroup outcomes under high threat because it shifts intergroup boundaries and, when there are high levels of threat, it creates a dissonance that is resolved by moving closer to the outgroup. This aligns with common ingroup identity models and existential threat theories (Pyszczynski et al. 2015). Future research would benefit from manipulating the context of threat or capturing naturally occurring contexts where threat varies (e.g., high threat, low threat, post-conflict) rather than measuring perceived threat as in the current research.

These findings further suggest that heightened threat amplifies sensitivity to interventions emphasizing shared humanity, aligning with theories that assert those who affiliate with more conservative ideologies are more attuned to threat-related cues (Onraet et al., 2014). The greater perceived threat in these groups underscores its role as a contextual amplifier, enhancing the salience of reconciliation efforts in polarized contexts (Borinca et al., 2024). These results refine meta-humanization theory by highlighting the contextual role of threat salience. Furthermore, the findings emphasize the importance of tailoring interventions to sociopolitical conditions, where perceived threat and emotional similarity become crucial drivers of reconciliatory outcomes (McDonald et al., 2017; Garrett & Bankert, 2018).

Future research has the potential to build upon our contributions by addressing limitations identified in our studies. First, in Studies 2a-3b we manipulated the focal predictor, however measured the chains of this process model correlationally. As a result, we cannot make definitive causal claims about whether moral emotions drive the pathway from meta-humanization to downstream outcomes. To establish causality, future research should manipulate these emotional responses (e.g., via moral emotion inductions, moral purity violations, or empathy-boosting interventions) and test whether changes in disgust, anger, or empathy alter the effects of meta-humanization on conciliatory attitudes. Such work would not only clarify the causal status of these mediators but also refine the theoretical model by determining whether moral emotions are necessary, sufficient, or merely facilitative components of the process. Importantly, our findings provide a theoretically grounded starting point for this work by identifying which emotions are most consistently implicated and under which conditions they operate.

Our research provided a comparative approach among Western bipartisan democracies. In line with this, the findings are specific to the sociopolitical contexts of the UK, US, and

Canada and may not generalize to other regions or cultures with different political dynamics, for example countries that do not have two-party systems or where political coalitions are commonplace. Second, we used the timely context of the, then upcoming, 54th US Presidential election to capture responses between American partisans, at a time when the potential threat to social cohesion should be heightened. However, the UK and Canadian studies were not implemented during comparable timepoints, for example before a looming election. The variation in levels of the moderating effect in the Democrat and Liberal samples likely reflects lower levels of perceived threat. When perceived threat to social cohesion is not salient, individuals may not feel as compelled to respond to methods promoting reconciliation (such as meta-humanization), as the stakes of intergroup conflict are less pertinent (Borinca et al., 2024). For purposes of replicability and ecological validity, it would be important for future research to be strategic and capture data before, during, and after critical social shifts like elections to demonstrate naturally occurring changes in perceived threat (see, for example, Roth et al., 2025).

In this research, we investigated threat to social cohesion as a moderator, potentially overlooking other important variables that could also play a moderating role, such as political efficacy, identity strength, or media exposure. Exploring these potential moderators is important for understanding meta-perceptions because these factors shape how individuals perceive how others view their group. For instance, exposure to partisan or biased media can exacerbate distorted perceptions, creating cycles of misjudgment about intergroup attitudes (Bail et al., 2018; Heseletine et al., 2024). Testing additional moderators will refine our understanding of how meta-perceptions develop and affect reconciliation efforts, providing a more thorough person x situation approach to the topic.

Our power analyses suggested that we had a sufficient sample to detect a small to medium effect size. We suggest that subsequent studies aim to gather a larger representative sample to further probe our findings related to empathy. Although there were inconsistent findings for empathy, the direct effects of meta-humanization to improve empathy were mostly significant and in the predicted direction across studies, as well as the direct effect of empathy to reduce affective polarization. This was expected as in intergroup contact contexts, empathy plays a crucial role in reducing stereotypes and intergroup anxiety, while improving trust and cooperation by addressing exaggerated beliefs about outgroup hostility (Dovidio et al., 2010; Vanman, 2016). Furthermore, empathy-driven interventions have proven effective in fostering political compromise, as demonstrated in a study on Brexit related polarization, where empathic perspective-taking reduced contempt and increased willingness to engage with opposing groups (Tausch et al., 2024), therefore we suggest more research is needed to tease apart these findings, specifically studies experimentally manipulating empathy.

An important limitation concerns the role of out-partisan contact, which emerged as a particularly strong predictor of conciliatory outcomes and, in some models, attenuated the effects of meta-humanization and moral emotions. This pattern is theoretically consistent with extensive evidence demonstrating the robust impact of intergroup contact on prejudice reduction (Paolini et al., 2021). When included as a covariate, contact likely absorbed substantial variance in outcomes, leaving less explanatory space for more nuanced meta-cognitive and affective mechanisms. Importantly, this does not undermine the present model but instead highlights that in contexts characterized by extensive or positive contact, experiential knowledge of the outgroup may supersede reliance on meta-perceptions when forming intergroup attitudes (Capozza et al., 2017). Recent research from Landry and Halperin (2025) documents the

motivational barriers to interventions addressing meta-dehumanization. They suggest that groups embroiled in conflict may not be motivated to engage with interventions aimed at ameliorating intergroup relations and that the positive effects meta-perceptions research is finding may only be applicable to the laboratory for certain intergroup dynamics (e.g., between Israelis and Palestinians). Applying this consideration to the findings from our studies, we acknowledge that interventions using meta-humanization are clearly not panacea. Although the benefits of meta-perception interventions may present boundaries for certain groups, we acknowledge that not all conflicts stem from the same roots, or involve the same dynamics. As such, future research should continue this line of investigation among varying group dynamics and different conflict settings in order to detail the profound benefits of meta-humanization and implement them into practical and impactful applications.

We suggest that the findings from this research could be used to develop interventions that incorporate collaborative problem-solving, where mixed groups of participants work together to draft resolutions that reflect shared values combining meta-perceptions with contact literature. This encourages the idea that, despite political divides, common ground can be found, facilitating cooperation and reconciliation. Previous studies have demonstrated that collaborative efforts across group lines can increase trust and cooperation (Borinca et al., 2024). Moreover, reflecting on past harmful rhetoric or actions could help participants recognize the moral consequences of division and foster accountability. Finally, interventions could conclude with feedback and reflection sessions where participants share how their perceptions of the opposing side have shifted. This reinforces the intervention's goal of reducing polarization and building mutual understanding.

## Conclusion

This research proposes a novel theoretical framework demonstrating that meta-humanization enhances cross-partisan attitudes through moral emotions. Across five studies, we demonstrate that perceiving one's political ingroup as humanized by the outgroup systematically reduces affective polarization and promotes conciliatory orientations by regulating moral emotional responses, most consistently by attenuating disgust. This shifts the focus from threat as the central explanatory mechanism of partisan hostility (Renström et al., 2023) to the moral-emotional processes that sustain symbolic exclusion and resistance to reconciliation.

A key contribution of this work is the identification of disgust as a critical emotional pathway linking meta-humanization to improved intergroup outcomes, highlighting its unique role in moral boundary maintenance and dehumanization in political conflict. By reducing disgust, meta-humanization appears to weaken the moralized perceptions that render outgroups as illegitimate, contaminating, or beyond engagement, thereby opening space for negotiation, cooperation, and contact.

Perceived threat to social cohesion functions as an important contextual boundary condition, shaping when these moral-emotional processes are most impactful. Rather than uniformly inhibiting reconciliation, heightened threat can amplify responsiveness to meta-humanization by regulating moral evaluations in polarized contexts. Together, these findings refine meta-humanization research by positioning moral emotions, particularly disgust, as central mechanisms through which humanizing meta-perceptions reduce political polarization.

By integrating meta-perceptions, moral emotions, and political partisanship within a unified framework, this research extends intergroup relations approaches into the domain of contemporary political conflict. In doing so, it offers a psychologically precise account of how

polarization can be mitigated, not by suppressing ideological disagreement, but by disrupting the moral-emotional processes that sustain partisan division.

## References

Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.B.F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media *115*(37), 9216-9221. https://doi.org/10.1073/pnas.1804840115

Bastian, B., & Haslam, N. (2011). Experiencing dehumanization: Cognitive and emotional effects of everyday dehumanization. *Basic and Applied Social Psychology*, *33*(4), 295–303. https://doi.org/10.1080/01973533.2011.614132

Berendt, J., van Leeuwen, E., & Uhrich, S. (2023). Can't live with them, can't live without them: The ambivalent effects of existential outgroup threat on helping behavior. *Personality and Social Psychology Bulletin, 50*(6), 971-984. https://doi.org/10.1177/01461672231158097

Borinca, I., Koc, Y. and Mustafa, S. (2024). Fostering social cohesion in post-conflict societies: The power of normative apologies in reducing competitive victimhood and enhancing reconciliation and intergroup negotiation. *European Journal of Social Psychology, 0*, 1-12 https://doi.org/10.1002/ejsp.3116

Boxell, L., Gentzkow, M., & Shapiro, J. (2024). Cross-country trends in affective polarization. *The Review of Economics and Statistics*, *106*(2), 557–565. https://doi.org/10.1162/rest_a_01160

Capozza, D., Di Bernardo, G.A., & Falvo, R. (2017). Intergroup contact and outgroup humanization: Is the causal relationship uni- or bidirectional? *PLoS One, 12*(1). https://doi.org/10.1371/journal.pone.0170554

Dovidio, J. F., Johnson, J. D., Gaertner, S. L., Pearson, A. R., Saguy, T., & Ashburn-Nardo, L. (2010). Empathy and intergroup relations. In M. Mikulincer & P. R. Shaver (Eds.), *Prosocial motives, emotions, and behavior: The better angels of our nature* (pp. 393–408). American Psychological Association. https://doi.org/10.1037/12061-020

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). How Affective polarization shapes Americans' political beliefs: A study of response to the COVID-19 Pandemic. *Journal of Experimental Political Science*, *8*(3), 223–234. https://doi.org/10.1017/XPS.2020.28

Engels, T., Traast, I.J., Doosje, B., Amodio, D., & Sauter, D. (2024). Moral elevation mitigates dehumanization of ethnic outgroups. *Current Research in Ecological and Social Psychology, 6*. https://doi.org/10.1016/j.cresp.2024.100187

Garrett, K.N. & Bankert, A. (2018). The moral roots of partisan division: How moral conviction heightens affective polarization. *British Journal of Political Science*, *50*(2):621-640. https://doi.org/0.1017/S000712341700059X

Giner-Sorolla, R., Martínez, R., Fernández, S., & Chas, A. (2023). Emotions as constituents, predictors and outcomes of dehumanization. *Current Opinion in Behavioral Sciences, 51*. https://doi.org/10.1016/j.cobeha.2023.101281

Giner-Sorolla, R., & Russell, P. S. (2019). Not just disgust: Fear and anger also relate to intergroup dehumanization. *Collabra: Psychology, 5*(1): 56. https://doi.org/10.1525/collabra.211

Golec de Zavala, A., Cichocka, A., & Iskra-Golec, I. (2013). Collective narcissism moderates the effect of in-group image threat on intergroup hostility. *Journal of Personality and Social Psychology, 104*(6), 1019–1039. https://doi.org/10.1037/a0032215

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Heseltine, M., Clemm von Hohenberg, B., Menchen-Trevino, E., Gackowski, T., & Wojcieszak, M. (2024). Effects of over-time exposure to partisan media and coverage of polarization on perceived polarization. *Political Communication*, 1–22. https://doi.org/10.1080/10584609.2024.2423093

Hess, U., & Landmann, H. (2018). Testing moral foundation theory: Are specific moral emotions elicited by specific moral transgressions? *Journal of Moral Education, 47*(1):34-47. https://doi.org/10.1080/03057240.2017.1350569

Jonas, E., McGregor, I., Klackl, J., Agroskin, D., Fritsche, I., Holbrook, C., Nash, K., Proulx, T., & Quirin, M. (2014). Threat and defense: From anxiety to approach. *Advances in Experimental Social Psychology, 49*, 219-286. https://doi.org/10.1016/B978-0-12-800052-6.00004-4

Kalisi, M. (2021). Relation between reliability towards social media with the scale of credibility on information on Kosovo-Serbia Dialogue. *International Journal of Soc. Hum. Sci*, *8*(15–16), 49–59

Kteily, N., Hodson, G., & Bruneau, E. (2016). They see us as less than human: meta-dehumanization predicts intergroup conflict via reciprocal dehumanization. *Journal of*

*Personality and Social Psychology, 110*, 343–370. http://dx.doi.org/10.1037/pspa0000044

Landry, A. P., & Halperin, E. (2025). Intergroup psychological interventions: The motivational challenge. *American Psychologist, 80*(2), 206–219. https://doi.org/10.1037/amp0001289

Landry, A.P., Ihm, E., & Schooler, J.W. (2022). Filthy animals: Integrating the behavioral immune system and disgust into a model of prophylactic dehumanization. *Evolutionary Psychological Science, 8*, 120–133. https://doi.org/10.1007/s40806-021-00296-8

Lee, A.H. (2022). Social trust in polarized times: How perceptions of political polarization affect Americans' trust in each other. *Political Behaviour, 44*(3),1533-1554. https://doi.org/10.1007/s11109-022-09787-1

McDonald, M., Porat, R., Yarkoney, A., Tagar, M. R., Kimel, S., Saguy, T., & Halperin, E. (2017). Intergroup emotional similarity reduces dehumanization and promotes conciliatory attitudes in prolonged conflict. *Group Processes & Intergroup Relations, 20*(1), 125–136. https://doi.org/10.1177/1368430215595107

Merkley, E. (2023). Polarization eh? Ideological divergence and partisan sorting in the Canadian mass public, *Public Opinion Quarterly*, *86*(4), 932-943 https://doi.org/10.1093/poq/nfac047

Moore-Berg, S.L., Ankori-Karlinsky, L.O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings National Academy of Science USA, 117*(26),14864-14872. https://doi.org/10.1073/pnas.2001263117

Moore-Berg, S. L., Hameiri, B., & Bruneau, E. G. (2022). Empathy, dehumanization, and misperceptions: A media intervention humanizes migrants and increases empathy for their plight but only if misinformation about migrants is also corrected. *Social Psychological and Personality Science, 13*(2), 645–655. https://doi.org/10.1177/19485506211012793

Onraet, E., Dhont, K., & Van Hiel, A. (2014). The relationships between internal and external threats and right-wing attitudes: A three-wave longitudinal study. *Personality and Social Psychology Bulletin*, *40*(6), 712-725. https://doi.org/10.1177/0146167214524256

Owuamalam, C. K., Tarrant, M., Farrow, C. V., & Zagefka, H. (2013). The effect of meta-stereotyping on judgements of higher-status outgroups when reciprocity and social image improvement motives collide. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement, 45,* 12–23. http://doi.org/10.1037/a0030012

Paolini, S., White, F.A., Tropp, L.R., et al. (2021). Intergroup contact research in the 21st century: Lessons learned and forward progress if we remain open. *Journal of Social Issues, 77*: 11–37. https://doi.org/10.1111/josi.12427

Parker, M.T., & Janoff-Bulman, R. (2013). Lessons from morality-based social identity: The power of outgroup "hate," not just ingroup "love". *Social Justice Research, 26*, 81–96. https://doi.org/10.1007/s11211-012-0175-6

Pew Research Center. (2022, August 9). *As Partisan Hostility Grows, Signs of Frustration With the Two-Party System*. https://www.pewresearch.org/politics/2022/08/09/as-partisan-hostility-grows-signs-of-frustration-with-the-two-party-system/

Prati, F., Crapolicchio, E., Dvorakova, A., Di Bernardo, G.A., & Ruzzante, D. (2023). Effective ways for reducing dehumanization: interpersonal and intergroup strategies. *Current Opinion in Behavioral Sciences, 51*. https://doi.org/10.1016/j.cobeha.2023.101277

Pyszczynski, T., Solomon, S., & Greenberg, J. (2015). Thirty years of terror management theory: From Genesis to Revelation. *Advances in Experimental Social Psychology, 52*, 1-70. https://doi.org/10.1016/bs.aesp.2015.03.001

Renström, E.A., Bäck, H., & Carroll, R. (2023), Threats, emotions, and affective polarization. *Political Psychology, 44*: 1337-1366. https://doi.org/10.1111/pops.12899

Roth, J., Steinmann, M., Loughnane, J., Devine, P., Muldoon, O., Shelly, C., ... & Campbell, C. (2025). Election results can decrease intergroup threat and through that positively affect intergroup relations. *Political Psychology, 46*(1), 7-23. https://doi.org/10.1111/pops.12960

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*(4), 574–586. https://doi.org/10.1037/0022-3514.76.4.574

Scatolon, A., Sharvit, K., Huici, C., Hernandez, A.A., Glazer, G., Sánchez, E.L., Michna, M. (2023). Focusing on the self to humanize others: The role of empathy and morality. *Current Opinion in Behavioral Sciences, 51*. https://doi.org/10.1016/j.cobeha.2023.101264

Schoemann, A.M., Boulton, A.J., Short, S.D., 2017. Determining power and sample size for simple and complex mediation models. *Soc. Psychol. Personal. Sci, 8*(4), 379–386. https://doi.org/10.1177/1948550617715068

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), Reducing prejudice and discrimination (pp. 225–246). Hillsdale, NJ: Lawrence Erlbaum

Tapias, M. P., Glaser, J., Keltner, D., Vasquez, K., & Wickens, T. (2007). Emotion and prejudice: specific emotions toward outgroups. *Group Processes & Intergroup Relations*, *10*(1), 27-39. https://doi.org/10.1177/1368430207071338

Tausch, N., Birtel, M.D., Górska, P., Bode, S., & Rocha, C. (2024). A post-Brexit intergroup contact intervention reduces affective polarization between Leavers and Remainers short-term. *Communications Psychology, 2*, (95). https://doi.org/10.1038/s44271-024-00146-w

Tybur, J. M., et al. (2020). Disgust, anger, and aggression: Further tests of the equivalence of moral emotions. *Collabra: Psychology, 6*(1): 34. https://doi.org/10.1525/collabra.349

van Doorn, J., Zeelenberg, M., & Breugelmans, S. M. (2014). Anger and prosocial behavior. *Emotion Review*, *6*(3), 261-268. https://doi.org/10.1177/1754073914523794

Vanman, E. J. (2016). The role of empathy in intergroup relations. *Current Opinion in Psychology, 11,* 59–63. https://doi.org/10.1016/j.copsyc.2016.06.007

van Zomeren, M., Postmes, T., & Spears, R. (2019). The motivational dynamics of collective action: Emotions as drivers of change. *Current Opinion in Psychology, 35*, 63-67. https://doi.org/10.1016/j.copsyc.2020.03.004

Varela, O., & Mead, E. (2018). Teamwork skill assessment: Development of a measure for academia. *Journal of Education for Business*, *93*(4), 172–182. https://doi.org/10.1080/08832323.2018.1433124

Vonhm, M. E. (2025). Key elements of social cohesion in conflict-affected societies: a critical literature review. *Cogent Social Sciences*, *11*(1). https://doi.org/10.1080/23311886.2025.2586176

Willis, J., Daniels, M., Disler, G., Khalil, L., & Zhou, A. (2017). Reliability and validity of the intergroup compromise inventory in two bipartisan samples. *SAGE Open*, *7*(4). https://doi.org/10.1177/2158244017739339

Table 1

*Main and interaction effects for dependent measures by political identification and meta-perception controlling for out-partisan contact in Study 1*

| | | ANCOVA | | | |
|---|---|---|---|---|---|
| Variable | df | Hostility *F* | Social distance *F* | Anger *F* | Disgust *F* |
| Political identification | 1, 349 | 1.69 | 8.90*** | 4.19* | 2.01 |
| Meta-perception | 1, 349 | 12.25*** | 6.03*** | 63.79*** | 97.18*** |
| Political identification x meta-perception | 1, 349 | 0.01 | 0.02 | 4.14*** | 1.72 |

*Note.* $p > .05$*, $p > .001$***.

Table 2

*Dependent measures as a function of political identification (Labour vs. Conservative) and meta-perception (meta-dehumanization vs. meta-humanization controlling for out-partisan contact) in Study 1*

| | | Hostility | | Social Distance | | Anger | | Disgust | |
|---|---|---|---|---|---|---|---|---|---|
| Political identification | Meta-Perception | M | SD | M | SD | M | SD | M | SD |
| Labour | Meta-Dehumanization | $2.55_{a}$ | 1.54 | $3.75_{a}$ | 0.94 | $2.61_{a,b}$ | 1.18 | $3.21_{a}$ | 1.32 |
| | Meta-Humanization | $2.05_{a}$ | 1.09 | $3.50_{a}$ | 0.95 | $1.94_{b}$ | 1.07 | $2.14_{a}$ | 1.23 |
| Conservative | Meta-Dehumanization | $2.32_{b}$ | 1.35 | $3.05_{b}$ | 1.03 | $2.77_{a,c}$ | 1.12 | $3.09_{b}$ | 1.18 |
| | Meta-Humanization | $1.82_{b}$ | 1.10 | $2.77_{b}$ | 0.96 | $1.64_{c}$ | 0.81 | $1.68_{b}$ | 0.95 |

*Note*. For meta-perception, columns sharing the same subscript are significantly different, p > .025

Table 3

*Unstandardized indirect effects (with standard errors) for Model 1 and Model 2 via moral emotions and humanization on intergroup outcomes, controlling for out-partisan contact in Study 2a*

| Predictor | Anger M1 | Disgust M2 | Empathy M3 | Humanization M4 | Affective Polarization (Y1) | Negotiation (Y2) | Compromise (Y3) | Teamwork (Y4) | Contact (Y5) |
|---|---|---|---|---|---|---|---|---|---|
| Model 1<br>MetaD – MetaH | -0.72 (0.23)*** | -0.82 (0.14)*** | 0.32 (0.24) | -0.39 (0.44) | 0.01 (0.14) | -0.004 (0.19) | -0.32 (0.17) | -0.23 (0.16) | -0.08 (0.16) |
| Model 2<br>Control - MetaH | -2.32 (0.23)*** | -1.13 (0.16) *** | 0.71 (0.28) ** | -0.90 (0.51) | 0.31 (0.17) | -0.25 (0.22) | -0.59 (0.20)*** | -0.45 (0.19)* | -0.37 (0.19) |
| Mediators | | | | | | | | | |
| Anger (M1) | | | | -0.24 (0.15) | 0.04 (0.05) | 0.03 (0.07) | -0.09 (0.06) | 0.04 (0.06) | 0.08 (0.06) |
| Disgust (M2) | | | | -0.47 (0.17) ** | 0.12 (0.06)* | -0.20 (0.08) ** | -0.10 (0.07) | -0.21 (0.07)*** | -0.17 (0.07) *** |
| Empathy (M3) | | | | 0.16 (0.11) | -0.17 (0.03)*** | 0.05 (0.05) | 0.10 (0.04)*** | 0.10 (0.04)** | 0.17 (0.04) *** |
| Humanization (M4) | | | | | -0.14 (0.02) *** | 0.15 (0.03) *** | 0.09 (0.02)*** | 0.08 (0.02)*** | 0.13 (0.02)*** |

Table 4

*Unstandardized indirect effects at 16th, 50th, and 84th percentiles of the moderator threat to social cohesion for Model 1 and Model 2 in Study 2a controlling for out-partisan contact*

| Model1 MD-MH | Affective Polarization | Negotiation | Compromise | Teamwork | Contact |
|---|---|---|---|---|---|
| Low (16th) | -.04 [-.10, .04] | .05 [-.03, .15] | .03 [-.02, .11] | .03 [-.02, .09] | .05 [-.04, .12] |
| Average (50th) | **-.06 [-.10, -.02]** | **.05 [.01, .11]** | **.03 [.01, .07]** | **.03 [.01, .06]** | **.05 [.01, .09]** |
| High (84th) | **-.06 [-.12, -.01]** | **.04 [.004, .11]** | .02 [-.01, .07] | .02 [-.001, .07] | **.03 [.003, .09]** |
| Model 2 Control-MH | | | | | |
| Low (16th) | -.05 [-.11, .04] | .06 [-.04, .17] | .03 [-.02, .12] | .03 [-.02, .10] | .05 [-.04, .14] |
| Average (50th) | **-.08 [-.14, -.03]** | **.07 [.02, .16]** | **.04 [.01, .10]** | **.04 [.01, .09]** | **.07 [.02, .13]** |
| High (84th) | **-.11 [-.23, -.04]** | **.08 [.01, .21]** | .04 [-.01, .14] | .05 [.00, .14 | **.06 [.01, .16]** |

Table 5

*Unstandardized indirect effects (with standard errors) for Model 1 and Model 2 via moral emotions and humanization on intergroup outcomes, controlling for out-partisan contact in Study 2b*

| Predictor | Anger M1 | Disgust M2 | Empathy M3 | Humanization M4 | Affective Polarization (Y1) | Negotiation (Y2) | Compromise (Y3) | Teamwork (Y4) | Contact (Y5) |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 MetaD – MetaH | -0.23 (0.20) | -0.56 (0.13)*** | 0.69 (0.22)** | 0.07 (0.41) | 0.28 (0.13)* | -0.01 (0.19) | -0.06 (0.14) | 0.03 (0.13) | -0.40 (0.17)** |
| Model 2 Control - MetaH | -2.97 (0.20)*** | -0.61 (0.17) *** | 1.57 (0.22) *** | 0.98 (0.43)* | 0.66 (0.17)*** | -0.45 (0.25) | -0.15 (0.18) | -0.33 (0.17) | -0.71 (0.23)*** |
| Mediators | | | | | | | | | |
| Anger (M1) | | | | 0.02 (0.19) | 0.13 (0.18)* | 0.02 (0.09) | 0.11 (0.06) | 0.06 (0.06) | -0.09 (0.08) |
| Disgust (M2) | | | | -0.44 (0.18)** | 0.04 (0.18) | -0.17 (0.08)* | -0.11 (0.06) | -0.16 (0.06)** | -0.01 (0.08) |
| Empathy (M3) | | | | 0.22 (0.10)* | -0.17 (0.10)*** | 0.02 (0.05) | 0.02 (0.03) | -0.02 (0.03) | 0.20 (0.04)*** |
| Humanization (M4) | | | | | -0.13 (0.02)*** | 0.09 (0.03)*** | 0.06 (0.02)*** | 0.03 (0.02) | 0.12 (0.03)*** |

Table 6

*Unstandardized indirect effects at 16th, 50th, and 84th percentiles of the moderator threat to social cohesion for Model 1 and Model 2 in Study 2b controlling out-partisan contact*

| Model1 MD-MH | Affective Polarization | Negotiation | Compromise | Teamwork | Contact |
|---|---|---|---|---|---|
| Low (16th) | -.04 [-.14, .02] | **.03 [.001, .07]** | **.02 [.002, .05]** | .01 [-.07, .11] | **.04 [.001, .09]** |
| Average (50th) | **-.03 [-.08, -.002]** | **.02 [.001, .06]** | **.01 [.001, .04]** | .03 [-.02, .10] | **.03 [.001, .07]** |
| High (84th) | -.02 [-.08, .01] | **.02 [.001, .06]** | .01 [-.001, .04] | .04 [-.04, .14] | .02 [-.002, .07] |
| Model 2 Control-MH | | | | | |
| Low (16th) | -.02 [-.06, .001] | -.02 [-.001, .05] | .01 [-.001, .03] | **.03 [.03, .07]** | .02 [-.001, .06] |
| Average (50th) | **-.04 [-.09, -.002]** | **.02 [.001, .07]** | **.02 [.001, .04]** | **.03 [.03, .07]** | **.03 [.001, .08]** |
| High (84th) | **-.04 [-.11, -.002]** | **.03 [.001, .09]** | **.03 [.001, .05]** | **.04 [.03, .08]** | **.04 [.001, .10]** |



Table 7

*Unstandardized indirect effects (with standard errors) for Model 1 and Model 2 via moral emotions and humanization on intergroup outcomes, controlling for out-partisan contact in Study 3a*

| Predictor | Anger M1 | Disgust M2 | Empathy M3 | Humanization M4 | Affective Polarization (Y1) | Negotiation (Y2) | Compromise (Y3) | Teamwork (Y4) | Contact (Y5) |
|---|---|---|---|---|---|---|---|---|---|
| Model 1<br>MetaD – MetaH | -0.17<br>(0.23) | -0.20<br>(0.14) | -0.04<br>(0.25) | -0.57 (0.42) | -0.003 (0.14) | 0.02 (0.20) | 0.03 (0.15) | 0.11 (0.18) | 0.12<br>(0.18) |
| Model 2<br>Control - MetaH | -0.91<br>(0.24)*** | -0.18<br>(0.15) | 0.64<br>(0.27)** | -0.48 (0.46) | 0.14 (0.15) | -0.19 (0.22) | 0.05 (0.16) | -0.10 (0.19) | 0.05<br>(0.19) |
| Mediators | | | | | | | | | |
| Anger (M1) | | | | -0.09 (0.20) | -0.004 (0.06) | 0.18 (0.10) | -0.02 (0.07) | 0.05 (0.09) | 0.06<br>(0.08) |
| Disgust (M2) | | | | -0.66<br>(0.20)*** | 0.16 (0.06)** | -0.25<br>(0.10)** | 0.04 (0.07) | -0.10 (0.09) | -0.18<br>(0.08)* |
| Empathy (M3) | | | | 0.12 (0.11) | -0.12<br>(0.04)*** | 0.11<br>(0.05)*** | 0.10 (0.04) ** | 0.04 (0.05) | 0.09<br>(0.05)* |
| Humanization (M4) | | | | | -0.14<br>(0.02)*** | 0.10<br>(0.03)*** | 0.10 (0.02)*** | 0.11<br>(0.03)*** | 0.16<br>(0.03)*** |

Table 8

*Unstandardized indirect effects at 16th, 50th, and 84th percentiles of the moderator threat to social cohesion for Model 1 and Model 2 in Study 3a controlling for out-partisan contact*

| Model1 MD-MH | Affective Polarization | Negotiation | Compromise | Teamwork | Contact |
|---|---|---|---|---|---|
| Low (16th) | .01 [-.03, .06] | -.01 [-.05, .03] | -.01 [-.04, .02] | -.01 [-.04, .03] | -.01 [ -.07, .04] |
| Average (50th) | -.02 [-.06, .01] | **.02 [.01, .05]** | **.02 [.01, .04]** | **.02 [.01, .06]** | **.03 [.01, .07]** |
| High (84th) | **-.05 [-.11, -.01]** | **.04 [.01, .08]** | **.04 [.01, .08]** | **.04 [.01, .08]** | **.06 [.01, .13]** |
| Model 2 Control-MH | | | | | |
| Low (16th) | .03 [-.01, .08] | -.02 [-.06, .01] | -.02 [-.06, .01] | -.02 [-.07, .01] | -.03 [-.09, .02] |
| Average (50th) | -.02 [-.06. .01] | .02 [-.01, .05] | **.02 [.01, .05]** | **.02 [.01, .05]** | .03 [-.01, .07] |
| High (84th) | **-.06 [-.14, -.02]** | **.04 [.01, .11]** | **.04 [.01, .10]** | **.05 [.01, .11]** | **.07 [.02, .16]** |

Table 9

*Unstandardized indirect effects (with standard errors) for Model 1 and Model 2 via moral emotions and humanization on intergroup outcomes, controlling for out-partisan contact in Study 3b*

| Predictor | Anger M1 | Disgust M2 | Empathy M3 | Humanization M4 | Affective Polarization (Y1) | Negotiation (Y2) | Compromise (Y3) | Teamwork (Y4) | Contact (Y5) |
|---|---|---|---|---|---|---|---|---|---|
| Model 1 MetaD – MetaH | -0.70 (0.24)*** | -0.49 (0.12)*** | 0.07 (0.23) | 0.26 (0.43) | 0.37 (0.14) | -0.46 (0.19) | 0.15 (0.15) | 0.13 (0.18) | -0.61 (0.16)** |
| Model 2 Control - MetaH | -1.16 (0.24)*** | -0.40 (0.13)*** | 0.66 (0.23)*** | 0.25 (0.43) | 0.39 (0.14) | -0.27 (0.19) | 0.03 (0.15) | 0.17 (0.18) | -0.42 (0.16)** |
| Mediators | | | | | | | | | |
| Anger (M1) | | | | 0.23 (0.21) | 0.04 (0.07) | 0.04 (0.09) | 0.06 (0.07) | 0.01 (0.09) | -0.01 (0.08) |
| Disgust (M2) | | | | -0.51 (0.21)* | 0.05 (0.07) | -0.12 (0.09) | -0.01 (0.07) | 0.06 (0.09) | -0.02 (0.08) |
| Empathy (M3) | | | | 0.30 (0.12)** | -0.23 (0.04)*** | 0.20 (0.5)*** | 0.13 (0.04)*** | 0.03 (0.05) | 0.28 (0.04)*** |
| Humanization (M4) | | | | | -0.16 (0.02)*** | 0.09 (0.03)*** | 0.09 (0.02)*** | 0.11 (0.03)*** | 0.14 (0.02)*** |

Table 10

*Unstandardized indirect effects at 16th, 50th, and 84th percentiles of the moderator threat to social cohesion for Model 1 and Model 2 in Study 3b controlling for out-partisan contact*

| Model1 MD-MH | Disgust → Affective Polarization | Empathy → Affective Polarization | Disgust → Negotiation | Empathy → Negotiation | Disgust→ Compromise | Empathy → Compromise | Disgust → Teamwork | Empathy → Teamwork | Disgust → Contact | Empathy → Contact |
|---|---|---|---|---|---|---|---|---|---|---|
| Low (16th) | **-.07 [-.14, -.01]** | -.02 [-.06, .01] | **.04 [.01, .09]** | .01 [-.01, .04] | **.04 [.01, .08]** | .01 [-.01, .03] | **.04 [.01, .09]** | .01 [-.01, .04] | **.06 [.01, 12]** | .01 [-.01, .05] |
| Average (50th) | **-.04 [-.08, -.01]** | .01 [-.03, .02] | **.02 [.01, .06]** | -.002 [-.01, .02] | **.01 [.01, .05]** | -.001 [-.01, .02] | **.03 [.01, .06]** | -.002 [-.02, .02] | **.03 [.01, .07]** | -.002 [-.02, .03] |
| High (84th) | -.01 [-.05, -.01] | .02 [-.02, .06] | .01 [-.01, .03] | -.01 [-.04, .01] | .001 [-.01, .03] | -.01 [-.03, .01] | .01 [-.01, .03] | -.01 [-.04, .01] | .01 [-.01, .04] | -.01 [-.06, .02] |
| Model 2 Control-MH | | | | | | | | | | |
| Low (16th) | **-.04 [-.10, -.01]** | **-.05 [-.11, -.01]** | **.02 [.01, .07]** | **.03 [.001, .07]** | **.02 [.003, .06]** | **.02 [.02, .07]** | **.03 [.01, .07]** | **.03 [.03, .08]** | **.04 [.01, .09]** | **.04 [.01, .10]** |
| Average (50th) | **-.03 [-.07. -.01]** | **-.03 [-.08, -.003]** | **.02 [.002, .05]** | **.02 [.001, .05]** | **.02 [.003, .04]** | **.02 [.01, .05]** | **.02 [.004, .05]** | **.02 [.002, .06]** | **.03 [.01, .06]** | **.03 [.002, .07]** |
| High (84th) | -.02 [-.06, -.01] | -.02 [-.07, .02] | .01 [-.003, .04] | .01 [-.01, .04] | .01 [-.003, .03] | .01 [-.01, .04] | **.01 [-.004, .04]** | .01 [-.01, .05] | .02 [-.01, .05] | .01 [-.01, .06] |

Supplementary Material

Study 2a ANCOVAS Republican Sample

Table 11

*Main effects for dependent measures by meta-perception controlling for out-partisan contact*

| | | ANCOVA | | | | |
|---|---|---|---|---|---|---|
| Variable | df | Affective polarization *F* | Negotiation *F* | Compromise *F* | Teamwork *F* | Contact *F* |
| Outgroup contact | 1, 302 | 42.34*** | 47.35*** | 47.79*** | 36.09*** | 75.83*** |
| Meta-perception | 2, 302 | 5.86** | 4.74** | 1.00 | 1.11 | 3.79* |

Table 12

*Dependent measures as a function of meta-perception (meta-dehumanization vs. meta-humanization vs. control) controlling for out-partisan contact*

| | Affective polarization | | Negotiation | | Compromise | | Teamwork | | Contact | |
|---|---|---|---|---|---|---|---|---|---|---|
| Meta-Perception | M | SD | M | SD | M | SD | M | SD | M | SD |
| Meta-Dehumanization | $4.74_{a,b}$ | 1.36 | $4.99_{a,b}$ | 1.73 | 5.06 | 1.38 | 5.70 | 1.29 | $5.70_{a}$ | 1.29 |
| Meta-Humanization | $4.09_{a}$ | 1.32 | $5.64_{a}$ | 1.20 | 5.34 | 1.14 | 5.98 | 0.98 | $5.98_{a}$ | 0.98 |
| Control | $4.38_{b}$ | 1.39 | $5.37_{b}$ | 1.47 | 5.07 | 1.32 | 5.80 | 1.28 | 5.80 | 1.28 |

Study 2b ANCOVAS Democrat Sample

Table 13

*Main effects for dependent measures by meta-perception controlling for out-partisan contact*

| | | ANCOVA | | | | |
|---|---|---|---|---|---|---|
| Variable | df | Affective polarization *F* | Negotiation *F* | Compromise *F* | Teamwork *F* | Contact *F* |
| Outgroup contact | 1, 302 | 21.28*** | 12.06*** | 16.65*** | 6.77** | 48.24*** |
| Meta-perception | 2, 302 | 2.66 | 0.51 | 0.10 | 0.89 | 1.49 |

Study 3a ANCOVAS Conservative Sample

Table 14

*Main effects for dependent measures by meta-perception controlling for out-partisan contact*

| | | ANCOVA | | | | |
|---|---|---|---|---|---|---|
| Variable | df | Affective polarization *F* | Negotiation *F* | Compromise *F* | Teamwork *F* | Contact *F* |
| Outgroup contact | 1, 302 | 3.81* | 4.74* | 0.15 | 0.69 | 11.80*** |
| Meta-perception | 2, 302 | 0.50 | 0.12 | 0.03 | 0.98 | 0.42 |

Study 3b ANCOVAS Liberal Sample

Table 15

*Main effects for dependent measures by meta-perception controlling for out-partisan contact*

| | | ANCOVA | | | | |
|---|---|---|---|---|---|---|
| Variable | df | Affective polarization *F* | Negotiation *F* | Compromise *F* | Teamwork *F* | Contact *F* |
| Outgroup contact | 1, 302 | 8.79*** | 19.39*** | 12.34*** | 9.21** | 27.16*** |
| Meta-perception | 2, 302 | 1.38 | 3.58* | 0.40 | 1.01 | 4.88** |

Table 16

*Dependent measures as a function of meta-perception (meta-dehumanization vs. meta-humanization vs. control)*

| | Affective polarization | | Negotiation | | Compromise | | Teamwork | | Contact | |
|---|---|---|---|---|---|---|---|---|---|---|
| Meta-Perception | M | SD | M | SD | M | SD | M | SD | M | SD |
| Meta-Dehumanization | 3.79 | 0.96 | 5.34 | 1.15 | 5.36 | 0.98 | 5.77 | 1.16 | $4.16_{a}$ | 1.20 |
| Meta-Humanization | 3.78 | 0.96 | $5.43_{a}$ | 1.29 | 5.49 | 1.00 | 5.94 | 1.08 | $4.83_{b}$ | 1.20 |
| Control | 4.09 | 1.18 | $4.86_{a}$ | 1.37 | 5.40 | 1.01 | 5.73 | 1.15 | $4.74_{a,b}$ | 1.29 |

Alternate Models

| Study | Alternate Model Description |
|---|---|
| Study 2a-3b (controlling for out-partisan contact) | Reverse Causality Model (meta-perceptions → conciliatory attitudes → humanization → disgust, with the path from meta-perceptions to conciliatory attitudes being moderated by threat to social cohesion) |
| | Disgust as Predictor Model (disgust → humanization → conciliatory attitudes) |



Two alternate models were tested to assess the robustness of the proposed ordering, both models again controlled for out-partisan contact. The reverse-causality model (meta-perceptions → conciliatory attitudes → humanization → disgust, with the path from meta-perceptions to conciliatory attitudes being moderated by threat to social cohesion) yielded no significant indirect effects. In contrast, an affect-driven model in which disgust served as the focal predictor, produced significant indirect effects, see Table 17. For example, higher disgust predicted lower outgroup humanization, which in turn was associated with increased affective polarization and reduced teamwork, negotiation, compromise, and desire for contact. This pattern aligns with extensive theorizing on disgust as a moral emotion that motivates exclusion and dehumanization, and underscores that moral emotions alone are insufficient to account for the positive associations seen in the initial proposed model on conciliatory outcomes. Collectively, the findings reinforced the hypothesized pathway and highlight the central role of meta-humanization in attenuating disgust and redirecting affective processes toward reconciliation.

Table 17

*Unstandardized indirect effect of disgust through humanization on dependent variables through out-partisan contact*

| Study | Affective Polarization | Negotiation | Compromise | Teamwork | Contact |
|---|---|---|---|---|---|
| 2a Republican | .11 [.07, .15] | -.09 [-.14, -.05] | -.06 [-.10, -.03] | -.08 [-.12, -.04] | -.09 [-.13, -.06] |
| 2b Democrat | .06 [.04. .09] | -.04 [-.08, -.01] | -.02 [-.04, -.004] | -.03 [-.09, -.02] | -.06 [-.09. -.03] |
| 3a Conservative | .11 [.06, .16] | -.07 [-.12, -.02] | -.07 [-.10, -.03] | -.07 [-.12, -.03] | -.12 [-.18, -.06] |

| | | | | | |
|---|---|---|---|---|---|
| 3b Liberal | .06 [.02, 10] | -.03 [-.07, -.01] | -.03 [-.06, -.01] | -.03 [-.06, -.01] | -.06 [-.10, -.02] |

R Studio Code for Power Analysis – Studies 2a-3b

```
#
==========
# POWER ANALYSIS FOR HYBRID PARALLEL-SERIAL MODERATED MEDIATION
#
==========
# Study design:
#   IV:         Meta-perception condition (meta-humanization vs. meta-dehumanization)
#   Moderator:   Perceived threat to social cohesion
#   Mediator 1:  Disgust     (parallel, moderated)
#   Mediator 2:  Anger       (parallel, moderated)
#   Mediator 3:  Empathy     (parallel, moderated)
#   Mediator 4:  Humanization (serial, not moderated)
#   DV:         Conciliatory attitudes
#
# Model structure:
#                 -> M1 (Disgust)  ->
#                 -> M2 (Anger)    -> M4 (Humanization) -> DV
#                 -> M3 (Empathy)  ->
#   IV (Threat moderates all IV -> M1/M2/M3 paths only)
#
# Three conditional indirect effects of interest:
#   1. IV -> Disgust  (mod. by Threat) -> Humanization -> DV
#   2. IV -> Anger    (mod. by Threat) -> Humanization -> DV
#   3. IV -> Empathy  (mod. by Threat) -> Humanization -> DV
```

```
#
# Coefficients used are the most conservative values across all four studies.
#
# HOW TO USE THIS SCRIPT:
#   1. Open RStudio
#   2. Paste this entire script into a new R Script file
#   3. If you haven't installed the packages yet, remove the #
#      from the install.packages lines below and run them first
#   4. Click Source to run the whole script
#   5. Results will appear in the Console panel
#
================================================================
===========




# ------------------------------------------------------------------------
# STEP 1: INSTALL AND LOAD REQUIRED PACKAGES
# ------------------------------------------------------------------------
# Remove the # from the lines below if you haven't installed these yet:
# install.packages("MASS")
# install.packages("lavaan")

library(MASS)
library(lavaan)



# ------------------------------------------------------------------------
# STEP 2: SET RANDOM SEED FOR REPRODUCIBILITY
```

```
# ------------------------------------------------------------------------

set.seed(123)



# ------------------------------------------------------------------------
# STEP 3: DEFINE MODEL PARAMETERS
# ------------------------------------------------------------------------

# --- IV -> Mediator paths (moderated by Threat) ---
a1_disgust  <- -0.84   # IV -> Disgust
a2_disgust  <-  0.06   # IV x Threat -> Disgust
a1_anger    <- -0.72   # IV -> Anger
a2_anger    <-  0.03   # IV x Threat -> Anger
a1_empathy  <-  0.31   # IV -> Empathy
a2_empathy  <-  0.03   # IV x Threat -> Empathy
a3          <-  0.20   # Threat main effect on all mediators

# --- Emotion -> Humanization paths (serial, not moderated) ---
d1 <-  0.24   # Disgust  -> Humanization
d2 <- -0.47   # Anger    -> Humanization
d3 <-  0.16   # Empathy  -> Humanization

# --- Humanization -> DV path ---
b4 <-  0.14   # Humanization -> DV



# ------------------------------------------------------------------------
```

```
# STEP 4: DEFINE SIMULATION SETTINGS
# ------------------------------------------------------------------------

n_simulations <- 5000   # Number of simulated datasets
sample_size   <- 300    # Most conservative sample size across studies
alpha         <- 0.05   # Significance threshold

# Moderator levels (Threat standardised: mean = 0, SD = 1)
threat_levels <- c(low = -1, medium = 0, high = 1)



# ------------------------------------------------------------------------
# STEP 5: DEFINE RESIDUAL VARIANCES
# ------------------------------------------------------------------------

compute_resid_var <- function(...) {
  coeffs <- c(...)
  v <- 1 - sum(coeffs^2)
  return(max(v, 0.05))
}

var_disgust      <- compute_resid_var(a1_disgust, a2_disgust, a3)
var_anger        <- compute_resid_var(a1_anger,   a2_anger,   a3)
var_empathy      <- compute_resid_var(a1_empathy, a2_empathy, a3)
var_humanization <- compute_resid_var(d1, d2, d3)
var_dv           <- compute_resid_var(b4)
```

```
# ------------------------------------------------------------------------
# STEP 6: RUN MONTE CARLO SIMULATION
# ------------------------------------------------------------------------

cat("============================================================\n")
cat("Running Monte Carlo power analysis...\n")
cat("5,000 simulations - this will take several minutes.\n")
cat("============================================================\n\n")

# Storage matrices: rows = simulations, columns = threat levels
ie_disgust <- matrix(NA, nrow = n_simulations, ncol = length(threat_levels),
                     dimnames = list(NULL, names(threat_levels)))
ie_anger   <- matrix(NA, nrow = n_simulations, ncol = length(threat_levels),
                     dimnames = list(NULL, names(threat_levels)))
ie_empathy <- matrix(NA, nrow = n_simulations, ncol = length(threat_levels),
                     dimnames = list(NULL, names(threat_levels)))

for (i in 1:n_simulations) {

  # --- Generate predictor variables ---
  iv          <- rbinom(sample_size, 1, 0.5)
  threat      <- rnorm(sample_size, mean = 0, sd = 1)
  iv_x_threat <- iv * threat

  # --- Generate parallel mediators (each independent of each other) ---
```

```
disgust <- a1_disgust * iv + a2_disgust * iv_x_threat + a3 * threat +
          rnorm(sample_size, 0, sqrt(var_disgust))

anger   <- a1_anger * iv + a2_anger * iv_x_threat + a3 * threat +
          rnorm(sample_size, 0, sqrt(var_anger))

empathy <- a1_empathy * iv + a2_empathy * iv_x_threat + a3 * threat +
          rnorm(sample_size, 0, sqrt(var_empathy))

# --- Generate serial mediator (humanization predicted by all three emotions) ---
humanization <- d1 * disgust + d2 * anger + d3 * empathy +
               rnorm(sample_size, 0, sqrt(var_humanization))

# --- Generate DV (predicted by humanization) ---
dv <- b4 * humanization +
     rnorm(sample_size, 0, sqrt(var_dv))

# --- Combine into data frame ---
sim_data <- data.frame(iv, threat, iv_x_threat,
                       disgust, anger, empathy,
                       humanization, dv)

# --- Fit model using lavaan ---
model_syntax <- '
  # Parallel mediator equations (all moderated by threat)
  disgust     ~ ad1*iv + ad2*iv_x_threat + ad3*threat
  anger       ~ aa1*iv + aa2*iv_x_threat + aa3*threat
```

```
  empathy      ~ ae1*iv + ae2*iv_x_threat + ae3*threat

  # Serial mediator equation (humanization predicted by emotions)
  humanization ~ dd1*disgust + dd2*anger + dd3*empathy

  # DV equation
  dv ~ b4*humanization
'

fit <- tryCatch(
  sem(model_syntax, data = sim_data),
  error = function(e) NULL
)

if (is.null(fit)) next

# --- Extract estimated coefficients ---
est <- lavaan::coef(fit)

ad1_e <- est["ad1"]; ad2_e <- est["ad2"]
aa1_e <- est["aa1"]; aa2_e <- est["aa2"]
ae1_e <- est["ae1"]; ae2_e <- est["ae2"]
dd1_e <- est["dd1"]; dd2_e <- est["dd2"]; dd3_e <- est["dd3"]
b4_e  <- est["b4"]

# --- Compute conditional indirect effects at each threat level ---
# Full indirect effect = (a1 + a2*Threat) * d * b4
```

```
  # where d is the path from each emotion to humanization

  for (j in seq_along(threat_levels)) {
    w <- threat_levels[j]

    ie_disgust[i, j] <- (ad1_e + ad2_e * w) * dd1_e * b4_e
    ie_anger[i, j]   <- (aa1_e + aa2_e * w) * dd2_e * b4_e
    ie_empathy[i, j] <- (ae1_e + ae2_e * w) * dd3_e * b4_e
  }
}



# ------------------------------------------------------------------------
# STEP 7: COMPUTE AND REPORT POWER
# ------------------------------------------------------------------------
# Expected directions:
#   Disgust:  (negative a1) * (positive d1) * (positive b4) = NEGATIVE
#   Anger:    (negative a1) * (negative d2) * (positive b4) = POSITIVE
#   Empathy:  (positive a1) * (positive d3) * (positive b4) = POSITIVE

cat("\n============================================================\
n")
cat("POWER ANALYSIS RESULTS\n")
cat("Hybrid Parallel-Serial Moderated Mediation\n")
cat("IV -> [Disgust/Anger/Empathy] (mod. Threat) -> Humanization -> DV\n")
cat("============================================================\n\
n")
```

```
cat("Model parameters used:\n")
cat(sprintf("  Disgust:       a1 = %.2f, a2 = %.2f, d1 = %.2f\n",
            a1_disgust, a2_disgust, d1))
cat(sprintf("  Anger:         a1 = %.2f, a2 = %.2f, d2 = %.2f\n",
            a1_anger, a2_anger, d2))
cat(sprintf("  Empathy:       a1 = %.2f, a2 = %.2f, d3 = %.2f\n",
            a1_empathy, a2_empathy, d3))
cat(sprintf("  Humanization -> DV (b4): %.2f\n", b4))
cat(sprintf("\nSample size:    %d\n", sample_size))
cat(sprintf("Simulations:    %d\n", n_simulations))
cat(sprintf("Alpha:          %.2f\n\n", alpha))

print_power_table <- function(ie_matrix, mediator_name, expected_positive = TRUE) {
  cat(sprintf("--- %s pathway ---\n", mediator_name))
  cat(sprintf("%-12s %-15s %-15s %-10s\n",
              "Threat Level", "Mean Indirect", "SD Indirect", "Est. Power"))
  cat(paste(rep("-", 55), collapse = ""), "\n")

  for (j in seq_along(threat_levels)) {
    effects   <- ie_matrix[, j]
    mean_ie   <- mean(effects, na.rm = TRUE)
    sd_ie     <- sd(effects, na.rm = TRUE)
    power_est <- if (expected_positive) {
      mean(effects > 0, na.rm = TRUE)
    } else {
      mean(effects < 0, na.rm = TRUE)
    }
```

```
    cat(sprintf("%-12s %-15.4f %-15.4f %-10.3f\n",
                names(threat_levels)[j], mean_ie, sd_ie, power_est))
  }
  cat("\n")
}

# Disgust:  negative * positive * positive = NEGATIVE indirect effect
print_power_table(ie_disgust, "DISGUST -> Humanization -> DV",
                  expected_positive = FALSE)

# Anger:    negative * negative * positive = POSITIVE indirect effect
print_power_table(ie_anger, "ANGER -> Humanization -> DV",
                  expected_positive = TRUE)

# Empathy:  positive * positive * positive = POSITIVE indirect effect
print_power_table(ie_empathy, "EMPATHY -> Humanization -> DV",
                  expected_positive = TRUE)

cat("------------------------------------------------------------\n")
cat("NOTE: Power is estimated as the proportion of simulations\n")
cat("where the conditional indirect effect was in the expected\n")
cat("direction for each pathway. This is a conservative estimate.\n")
cat("------------------------------------------------------------\n\n")

cat("INTERPRETATION GUIDE:\n")
cat("  Power >= 0.80 = Adequately powered\n")
cat("  Power >= 0.90 = Well powered\n")
```

```
cat("  Power <  0.80 = Underpowered\n\n")



============================================================
POWER ANALYSIS RESULTS
Hybrid Parallel-Serial Moderated Mediation
IV -> [Disgust/Anger/Empathy] (mod. Threat) -> Humanization -> DV
============================================================

Model parameters used:
  Disgust:       a1 = -0.84, a2 = 0.06, d1 = 0.24
  Anger:         a1 = -0.72, a2 = 0.03, d2 = -0.47
  Empathy:       a1 = 0.31, a2 = 0.03, d3 = 0.16
  Humanization -> DV (b4): 0.14

Sample size:    300
Simulations:    5000
Alpha:          0.05

--- DISGUST -> Humanization -> DV pathway ---
Threat Level Mean Indirect   SD Indirect     Est. Power
-------------------------------------------------
low                    -0.0301         0.0171          0.989
medium                 -0.0282         0.0159          0.989
high                   -0.0262         0.0149          0.989

--- ANGER -> Humanization -> DV pathway ---
```

Threat Level Mean Indirect SD Indirect Est. Power

---

low 0.0491 0.0242 0.991

medium 0.0471 0.0225 0.991

high 0.0450 0.0222 0.991

--- EMPATHY -> Humanization -> DV pathway ---

Threat Level Mean Indirect SD Indirect Est. Power

---

low 0.0063 0.0052 0.960

medium 0.0070 0.0047 0.989

high 0.0076 0.0057 0.977

---

NOTE: Power is estimated as the proportion of simulations where the conditional indirect effect was in the expected direction for each pathway. This is a conservative estimate.

---

INTERPRETATION GUIDE:

Power >= 0.80 = Adequately powered

Power >= 0.90 = Well powered

Power < 0.80 = Underpowered