



OPEN ACCESS

EDITED BY

Jazmin Brown-Iannuzzi,
University of Virginia, United States

REVIEWED BY

Jin Xun Goh,
Colby College, United States
Michael Pasek,
University of Illinois Chicago, United States

*CORRESPONDENCE

Kevin R. Carriere
✉ krcarriere@gmail.com

†PRESENT ADDRESS

Rebecca Ghion,
Department of Sciences, John Jay College for
Criminal Justice, City University of New York,
New York, NY, United States

RECEIVED 12 January 2024

ACCEPTED 07 March 2024

PUBLISHED 28 March 2024

CITATION

Carriere KR, Hallenbeck G, Sullivan D and
Ghion R (2024) You are not like the rest of
them: disrupting meta-perceptions dilutes
dehumanization.
Front. Soc. Psychol. 2:1369709.
doi: 10.3389/frsps.2024.1369709

COPYRIGHT

© 2024 Carriere, Hallenbeck, Sullivan and
Ghion. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

You are not like the rest of them: disrupting meta-perceptions dilutes dehumanization

Kevin R. Carriere^{1*}, Gregory Hallenbeck², Delaney Sullivan³ and
Rebecca Ghion^{3†}

¹Department of Psychology, Stonehill College, Easton, MA, United States, ²Department of Mathematics and Statistics, Binghamton University, Binghamton, NY, United States, ³Department of Psychology, Washington and Jefferson College, Washington, PA, United States

Political polarization, fueled by conflicting meta-perceptions, presents a critical obstacle to constructive discourse and collaboration. These meta-perceptions—how one group perceives another group's views of them—are often inaccurate and can lead to detrimental outcomes such as increased hostility and dehumanization. Across two studies, we introduce and experimentally test a novel approach that exposes participants to atypical, counter-stereotypical members of an opposing group who either confirm or disrupt their existing meta-perceptions. We find that disrupting meta-perceptions decreases dehumanization of the partner, increases interest in wanting to learn more about them, but fails to increase willingness to interact in the future with the partner. We conduct an exploratory text analysis to uncover differences in word choice by condition. Our research adds a new dimension to the existing body of work by examining the efficacy of alternative intervention strategies to improve intergroup relations in politically polarized settings.

KEYWORDS

meta-perceptions, imagined interactions, text analysis, meta-perception correction, dehumanization, atypicality, counter-stereotypical exemplars

Introduction

Political conflict is high, and calls to reconcile with our political counterparts seem more of a fever dream by the day. Among American voters, 59% say that they cannot agree on basic facts with their political opponents, never mind differences in policies (Connaughton, 2021). Beyond this, 69% of Republicans and 83% of Democrats say that their counterparts are close-minded (Pew Research Center, 2022). Much of this conflict is built around meta-perceptions—what we think the other thinks of us. Studies have shown that these meta-perceptions are often exaggerated and wildly inaccurate (Lees and Cikara, 2020; Mernyk et al., 2022). In two experimental studies, we investigate whether disrupting participants' meta-perceptions can decrease dehumanization toward an individual member of an out-group, foster a curiosity to learn more about the other party, and boost intentions for future interactions with the individual. Study 1 explores a potential boundary condition on how the medium (voice or text) in which this information is conveyed might affect perceptions, while Study 2 examines more directly interest in future interactions. We also employ text analysis to better understand the language patterns prevalent in these interactions as a new and fruitful descriptive analysis of political conversations.

Defining meta-perceptions

Social identity theory (Tajfel and Turner, 1979) posits that groups are created and maintained through a three-step process of categorization, identification, and comparison, by which individuals are constantly considering the positions, values, and beliefs of groups; choosing which they identify most closely with; and then comparing one's newly identified group with the positions, values, and beliefs of other groups. Underlying this process is the understanding that all individuals are engaging in this process at all times and that we are aware of the fact that others are engaging in categorizations and comparisons as well (while we are unable to know what these meaning comparisons exist as). In doing so, by identifying with a group in this way, we must hold (a) our self-perceptions, that is, what we think of our group; (b) perceptions, what we think of the other group; and (c) our meta-perceptions, what we think the other group is thinking of our own group. Meta-perception research has primarily focused on either meta-dehumanization (believing that the other sees your group as less than human) or meta-prejudice (believing that the other dislikes your group).

It is critical that the concept of meta-perceptions is elaborated through this dialogic relationship of individual to other to individual in order to accurately grasp the psychological process at play. This can be in terms of how one imagines the opposing political party feels toward your party on feeling thermometers or on how much humanity the other group would attribute to one's own party (Vorauer et al., 1998; Moore-Berg et al., 2020; Landry et al., 2021). This is in contrast to other work that operationalizes meta-perceptions as the difference between how much one believes their in-group values something compared to how much they perceive the out-group values that same characteristic (Pasek et al., 2022) or estimating how another group may act in a given situation (Shackelford et al., 2024). Others use meta-perceptions as in-group social norm perceptions, such as asking certain group members (international students, individuals with mental health problems, and Muslims) to what extent they think that members of their group would want to do certain actions, such as the desire for contact with other groups (Stathi et al., 2020) or rating of out-group actions (Lees and Cikara, 2020). This body of work, while insightful in understanding out-group social norm perceptions, fails to capture the *meta* part—perceiving how another perceives *your* own actions and beliefs—of the construct (Moore-Berg et al., 2020).

This self-focus is particularly important since meta-perceptions of one's own group can lead to serious intergroup consequences such as increased hostilities toward that group (O'Brien et al., 2018), increased dehumanization and stereotyping of the out-group (Moore-Berg et al., 2020), and increases in adherence to your own political party (Ahler and Sood, 2018). Participants who were led to believe that they were communicating with a partner of a different race showed that the more they thought their partner held negative stereotypes of them, the higher intergroup anxiety they reported experiencing (Finchilescu, 2010), and the higher aversion they had to future interactions with that group (Vorauer et al., 1998). Meta-perceptions have also been linked to increased negative perceptions of the other group (Alexander et al., 2005), obstructionism (Lees

and Cikara, 2020), and increased intergroup hostility (Moore-Berg et al., 2020).

Across 10 samples, Kteily et al. (2016) showed that increases in meta-dehumanization were related to increases in support for rights violations (support for torture, opposition to immigration, aggressive policies) through increases in dehumanization of the target group. This action of dehumanization—of viewing another as less than human, either through considerations of evolution, relating another to animals, or stripping away their human nature—has been found to be independent of stereotyping and discrimination (Kteily et al., 2015; Landry et al., 2022). Believing the other side was dehumanizing the participant led to a direct effect increase in support for civil liberty violations as well as indirectly through increases in participants' own dehumanization of the targets. Just as violence begets more violence, research has consistently shown how perceived dehumanization begets more dehumanization (Kteily et al., 2016; Kteily and Bruneau, 2017; Stathi et al., 2020; Bruneau et al., 2021).

In the research reviewed, meta-perceptions generally are operationalized as how the other (de)humanizes one's group (Kteily et al., 2016; Kteily and Bruneau, 2017; Moore-Berg et al., 2020; Landry et al., 2022, 2023; Petsko and Kteily, 2023; meta-dehumanization) or how the other feels toward one's group (Vorauer et al., 2000; Finchilescu, 2010; Kteily et al., 2016; Kteily and Bruneau, 2017; Moore-Berg et al., 2020; Landry et al., 2022, 2023; meta-prejudice). However, there are more than two ways to inaccurately judge the beliefs of one's group. One may misperceive how one is judged in terms of intelligence, skill, or in views of the world. For example, in one study, individuals of both political parties categorized ambiguous faces into political parties based on what they believed an opposing political party member would report (Petsko and Kteily, 2023). Other studies have examined the meta-perception of support for violence from a political opponent (Mernyk et al., 2022) or how the other views oneself generally (O'Brien et al., 2018).

While meta-dehumanization and meta-prejudice have been shown to be critically important to intergroup processes, they are not the sole meta-perceptions that can influence group relations. Therefore, a current limitation in the literature that we seek to address is the need to expand the conceptualization of a meta-perception beyond meta-dehumanization and meta-prejudice toward everyday political beliefs.

Meta-perception interventions

Much of the focus of research has been on the accuracy of these perceptions of future social interactions. To assess this, researchers ask participants of group X to predict how out-group Y reports feelings toward group X (the meta-perception under investigation). They also seek out out-group Y and get the overall average of that group's reported feelings toward group X (Saguy and Kteily, 2011). A difference between these two numbers would suggest a difference between what is being perceived and what the underlying truth is. In a global study of more than 10,000 participants and 26 countries, there was a clear effect that meta-perceptions are vastly overestimated compared to the actual perception (Ruggeri et al., 2021).

Because the primary focus of research has been on improving the accuracy of meta-perceptions, a common intervention strategy generally entails informing participants of the true perceptions of the opposing group, known as the “misperception correction.” For example, [Lees and Cikara \(2020\)](#) showed that meta-perceptions of how much a political opponent would dislike an action by one’s party predicted the belief that the out-group was motivated by obstructionism. In a later experiment, they used data from the prior experiment to inform participants about their political opponent’s opinions on how much they disliked a certain topic (which was generally lower than the meta-perception had predicted). They found that the larger the difference in meta-perceptions, the more effective the intervention was at reducing obstructionism ratings, a finding replicated at a global scale ([Ruggeri et al., 2021](#)). Others have found that this misperception correction leads participants to report more moderate policy positions themselves ([Ahler, 2014](#)) and be more supportive of democratic norms toward the opposing political party ([Landry et al., 2023](#)). Importantly, time and time again, these interventions have shown success in reducing dehumanization of the opposing side ([Lees and Cikara, 2020](#); [Moore-Berg et al., 2020](#); [Ruggeri et al., 2021](#); [Landry et al., 2023](#)).

While these misperception interventions reduce affective polarization, they do little to reduce support for anti-democratic policies performed by members of one’s own political group ([Broockman et al., 2023](#); [Voelkel et al., 2023](#)), and the interventions did not survive a 1-week follow-up ([Lees and Cikara, 2020](#)). A more recent vignette-styled meta-perception correction treatment effect survived a 1-week follow-up in terms of reducing dehumanization and prejudice but did not show durable support for democratic policies ([Landry et al., 2023](#)). These results do stand in contradiction to a recent study that showed lasting effects 1 month later ([Mernyk et al., 2022](#)), when participants who were informed of the actual perception of out-group members showed significantly less willingness to engage in violence and lowered the perception of the out-group member’s willingness to engage in violence.

Boundary conditions for reducing dehumanization

Meta-perception correction interventions, as reviewed, show durable results in decreasing dehumanization ([Lees and Cikara, 2020](#); [Moore-Berg et al., 2020](#); [Ruggeri et al., 2021](#); [Landry et al., 2023](#)). Of course, other interventions to improve intergroup relations and decrease dehumanization external to targeting meta-perceptions exist, and our review does not discount their presence. Indeed, interventions such as imagined contact ([Yetkili et al., 2018](#)), actual intergroup contact ([Zingora et al., 2021](#)), and intergroup cooperation ([Nomikos, 2022](#)) are all ways in which current work has attempted to reduce intergroup aggression.

It is not necessarily the case that all interventions are equal, or that some interventions may not be better than others in certain circumstances. For example, meta-perceptions are generally larger among stronger partisans ([Pasek et al., 2022](#)) and potentially in particular Democrats ([Landry et al., 2023](#), but cf. [Lees and Cikara, 2020](#), for no difference in party), and correction interventions work best for individuals who are highly inaccurate compared to those who are only slightly inaccurate ([Mernyk et al., 2022](#)).

Meta-perception inaccuracy is reduced when one perceives their group to be facing significant losses ([Saguy and Kteily, 2011](#)) or when individuals perceive that the out-group will be highly upset about the policy ([Lees and Cikara, 2020](#)). By comparison, meta-perceptions of an out-group are not moderated by how much a group wishes for future interactions ([O’Brien et al., 2018](#)).

However, the tangible benefit from measuring accuracy in meta-perceptions seems at best theoretical. Individuals are not aware of their accuracy and can rarely find out what the out-group truly believes, external to the invisible hand of the researcher informing them of their inaccuracy. Worse, polls that may explain the inaccuracy of those beliefs may be easily deemed ‘fake news’ by others and dismissed if they do not conform to one’s worldview ([Harper and Baguley, 2022](#)).

Thus, we posit that interventions styled in the terms of “Well, actually (this out-group) rated your group as this percent human” will do little to truly alleviate the situation and that new interventions need to be developed. Due to the strong correlation of meta-perceptions with dehumanization, we could consider non-meta-perception interventions that have sought to reduce dehumanization. For example, [Schroeder et al. \(2017\)](#) showed that in online studies, using voice, rather than text, when reading about an individual’s political beliefs resulted in less dehumanization of the opponent. As a relatively straightforward manipulation, and with a large reported effect size (*ds* range from .29 to .73), this intervention serves as a good boundary condition for interventions surrounding dehumanization and meta-perceptions. We test this directly in Study 1.

Beyond this, instead of presenting information that can easily be dismissed, presenting individuals with an exemplar of such a misperception may be more externally valid. If we can have contact someone who disrupts our meta-perception, we may be able to alleviate the consequences of inaccurate meta-perceptions, such as increased dehumanization and an unwillingness to engage in future interactions with an individual from the out-group. Next, we review the research around atypicality as a potential intervention to engage with meta-perceptions.

Atypicality as an intervention

In their seminal work discussing the application of meta-perceptions to an intergroup context, [Frey and Tropp \(2006\)](#) note how when considering meta-perceptions, individuals expect to be viewed in terms of the stereotype of their group when being examined by an out-group member ([Vorauer et al., 2000](#)). If we wish to decrease both the causes and effects of meta-perceptions, then turning our attention toward stereotypical thinking—and its converse—may be particularly effective, given that thinking about counter-stereotypical individuals increases the humanization of not only those individuals but also of various unrelated out-group members ([Prati et al., 2015](#)).

There is evidence to suggest that typicality interventions could help reduce conflict surrounding meta-perceptions, with a meta-analysis showing a medium-sized effect ($g = .58$) that out-group member information is generalized from the individual toward the group, particularly under conditions of moderate atypicality in comparison to extreme atypicality ([McIntyre et al., 2016](#)). In

one study, atypical members of a Border and Custom Agents Association were rated as less threatening and more favorable than typical out-group members (Yetkili et al., 2018), and interacting with atypical exemplars suffering from mental illness showed decreases in stigma around mental health (Maunder and White, 2023). In examining a race-based implicit attitudes test, participants who were manipulated to think about Black leaders, in comparison to White leaders, showed no implicit racial bias against Black individuals, and this effect lasted even after a 24-h delay in measurement (Dasgupta and Greenwald, 2001). While a recent replication failed to directly replicate this earlier study, it did show that atypicality could reduce bias toward neutrality if participants recognized the atypical link being shown (Kurdi et al., 2023). Finally, a benefit of atypicality interventions is in their flexibility over the target group. Atypicality interventions work for not just out-group member interactions but also one's own viewing of their own in-group (Lai et al., 2014). In one study, women who were exposed to women in leadership roles (atypical) were less biased in a proceeding implicit attitudes test measuring bias against women in leadership compared to women exposed to a control condition (Dasgupta and Asgari, 2004).

Including counter-stereotypical exemplars in interactions may reduce the tendency to rely on stereotypical assumptions (Prati et al., 2015; Hodson et al., 2018). In one study, Czech individuals who believed that they would have negative contact with a disliked out-group (Roma individuals) were the ones who most benefited from experiencing positive contact (stereotype-inconsistent) with Roma individuals (Zingora et al., 2021). Overgaard et al. (2021) argue that the state of polarization across the globe, in large part due to meta-perceptions, requires a renewed focus on building interactions that are tailored to a 'connective democracy'. They propose that we must find areas of commonality and likeness to truly bridge the political divide. We believe this starts first with contact—and contact of a singular person. If we can improve interpersonal interactions, it provides an opportunity to then use these positive intergroup contact moments as a starting place to improve intergroup relations. While this is an indirect method to target intergroup attitudes, explicitly addressing broader group attitudes is beyond the scope of this article.

Methodological advancement in understanding perceptions

Finally, we must continue to press the investigation of external validity and theory building. There has been a recent movement to reconsider the value of descriptive data (Hofmann and Grigoryan, 2023; Bonetto et al., 2024). While predictive analyses have their value, descriptive analyses can provide deeper knowledge about psychological phenomena and can provide generative platforms for new hypotheses and experiments for quantitative work to explore (Power et al., 2018). When studying something as applicable, complex, and emotionally salient as intergroup relations within politics, it may be particularly important to ensure we clearly understand the phenomenon under examination. The limitations of reported attitudes predicting behavior have been well documented, and there has been a concern over the decrease in

behavioral observations in leading social psychology journals for some time (Baumeister et al., 2007; Simon and Wilder, 2022).

With the advent of social media and “large” data sets, psychology can apply other methods to previously under-analyzed data sets, such as language use. In one study, Kubin et al. (2021) analyzed emotional tone scores on YouTube videos that were framed around personal experiences with abortion or the facts about abortion. They showed a clear difference between the kind of language used when the video discussed personal experiences; the comments left were more positive in their tone and word choice. They used the Linguistic Inquiry and Wordcount Software (LIWC) for their text analysis to quantify what percentage of words in a comment were in various emotion-based categories (Pennebaker et al., 2003). Another study examined writings that promote group hostility and found that more words relate to disgust and hate than one would expect in general language use (Taylor, 2007). These examples show the benefit of observation and analysis of word use in helping theory move forward (Bonetto et al., 2024; Mulderig et al., 2024). Current interventions of meta-perceptions have revolved around the pre-/post measurement of self-reported attitudes and beliefs. While useful in many ways – and we ourselves undertake similar analyses – it leaves much to be desired in terms of asking if these interventions are having a tangible, qualitative impact on the participants themselves.

Therefore, we highlight three current limitations with the published research on meta-perceptions. First, the current intervention of informing individuals that ‘the truth’ out in the world is not how most people consume information and that such information could be easily dismissed among highly motivated individuals. To alleviate this concern, we hypothesize that presenting an individual with an exemplar that disproves one's meta-perception may hold more external validity in considering how one may combat these misperceptions. In this regard, we also turn our focus to addressing interpersonal interactions with the hope that it will translate to improved intergroup interactions (McIntyre et al., 2016). Second, we note that current research has been focused primarily on targeting meta-dehumanization and meta-prejudice. While extremely important based on their predictive power and correlations with other constructs of interest, there exist other forms of perceptions that are equally as deserving of attention and investigation, and more investigation into other outcomes of meta-perception correction interventions, such as interest in intergroup contact, are needed. Finally, we note a need to expand the methods used to analyze and engage with interventions and a need to begin the process of putting behavior at the forefront of social psychological research.

In this article, we experimentally test a novel meta-perception intergroup interaction by exposing participants to partners who either confirm or disrupt their highest personally held meta-perception and then learn more about their partner. What they learn about their partner is designed as a communicative (if not cooperative) framework of interaction based on research that supports a cooperative framework (Eisenkraft et al., 2017). In Study 1, we examine the moderating effect of how participants interact with their partner (either through text or voice), due to research suggesting that this may be an important consideration in dehumanization (Schroeder et al., 2017). Finding no significance,

TABLE 1 List of possible meta-perceptions.

Republicans think all democrats	Democrats think all republicans
Want to steal our guns.	Want to give every American a gun.
Want to open all borders and give citizenship to all undocumented individuals.	Want to close all borders and deport all DACA recipients.
Want everyone to be able to marry everyone and everything, including their dog.	Want to ban all support and protections for LGBTQI+ individuals.
Want abortion to be so freely accessible it becomes a normal form of birth control.	Want abortion to be illegal in all circumstances, including rape and incest, and jail individuals who have them.
Want to remove all funding from the military.	Want to seriously reduce funding from social services in order to spend more on the military.
Want to remove all funding for the police.	Want to seriously reduce funding for social services in order to spend more on the police.
Support the Black Lives Matter movement and Black racial domination.	Are against the Black Lives Matter movement and support white racial domination.

DACA, Deferred Action for Childhood Arrivals.

we remove this moderator in Study 2 and measure future interaction intentions more directly. Finally, in an exploratory analysis spanning both Studies 1 and 2, we use text analysis to see if the interactions with their partner can predict one's willingness to interact in the future. All materials are hosted on the Open Science Framework (OSF; https://osf.io/xqrtn/?view_only=c3ac648283824972be0c37449effe2b9). Based on the prior research, we hypothesize that disrupting a meta-perception should lead to decreased dehumanization of the out-group member, increased curiosity to learn more about the out-group member, and a willingness to interact with an individual of the opposing political party.

Study 1

Participants and procedure

The sample consisted of 373 participants from an online subject pool platform (Luc.id; $M_{age} = 47.88$, $SD_{age} = 17.02$; Republican = 38.42%, Democrat = 61.58%; male = 37.32%, female = 61.85%, non-binary = 0.8%). Participants were removed if they (a) did not vote in the 2020 election, (b) their vote did not match their party, or (c) they voted for a ticket that was not Republican or Democrat. We also collected their race, ethnicity, first name or pseudonym, and their dehumanization of the opposing political party. The dehumanization of the opposing political party measure was for a separate study and was not within our hypothesized pathway, so we do not analyze these measures further.

After, participants were presented with seven possible meta-perceptions of what the opposing party thinks of their party and asked to rank-order them from the top ("I believe this

the most") to the bottom ("I believe this the least"; see Table 1 for a full list). That is, Democrats would be asked to rank the first column of Table 1, while Republicans would rank the second column. The meta-perceptions were designed to be extreme in order to reflect the general inaccuracy of meta-perceptions.

After rank-ordering their meta-perceptions, participants engaged in a filler task in which they rated their own perceptions of the opposing meta-perceptions on a 7-point scale from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*; a Democrat was then asked "How much do you agree that Republicans want to give every American a gun?"). This task functions as a way to make the following manipulation more believable and is not used in any analysis.

After completing the filler task, participants were informed they were going to need to explain why they voted for their political candidate. But before they did that, they would be randomly paired with a prior participant who opted to share their response with the participant and opted to receive any response they might decide to send about why they had voted for their candidate of choice. After clicking continue, participants waited 10 seconds while the system matched them with 'Casey', a gender-neutral name.

Through survey logic, Casey was the same age, belonged to the opposing political party, and voted for the opposing political candidate compared to the participant. Through random assignment, they were told that Casey had either strongly agreed (confirming meta-perception) or strongly disagreed (disrupting meta-perception) with the perception that was the participant's highest-ranked meta-perception. Framed in this way, Casey appears to be a prior participant who had completed the filler task and either confirmed or disrupted their highest-ranked meta-perception. For example, a 25-year-old Democrat participant who ranked that Republicans believe that Democrats want to steal their guns and was assigned to be in the Disrupted condition would be presented with the following text: "Your partner, Casey, is 25 years old and is a Republican who voted for Donald Trump and Mike Pence. In their survey, they reported that they strongly disagree with the statement "Democrats want to steal our guns."

We reminded them again of the description of Casey (age, party, vote choice, and manipulation) and collected participants' interest in understanding the perspective of Casey and their subtle dehumanization of Casey. Participants then were randomly presented to either listen to or read the reasons why Casey voted for the opposing candidate¹ (voice gender was randomly assigned), and those who passed an attention check for why Casey stated to have voted for the party were given the opportunity to write a message back to Casey. Participants then reported the persuasiveness of the story and again filled out a post-reading subtle dehumanization of Casey measurement. They were then thanked and debriefed.

1 Full information regarding the development and stimuli of Casey's reasons can be found on the OSF under Pilot Study (https://osf.io/xqrtn/?view_only=c3ac648283824972be0c37449effe2b9).

Measures

Interest in understanding perspective

Participants responded to four questions about their interest in reading Casey's reasoning on a 7-point scale from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*; I am willing to learn about Casey's reasons for why they voted for __; I am excited to learn about Casey's reasoning for why they voted for __; I do not want to learn about Casey's reasoning for why they voted for __ (reverse-coded); I do not care about Casey's reasoning for why they voted for __ (reverse-coded); $\alpha = 0.87$, $M = 4.27$, $SD = 1.7$).

Subtle dehumanization of Casey

Participants were presented with a list of 12 traits and were asked to rate from 1 (*Much less than the average person*) to 7 (*Much more than the average person*) how much they believed Casey exhibited these traits. These 12 traits come from prior work measuring subtle dehumanization and include traits such as open-minded (reverse-coded) and unsophisticated (Bastian and Haslam, 2010). This measure was presented both before ($\alpha_{pre} = 0.81$, $M_{pre} = 3.92$, $SD_{pre} = 0.84$) and after ($\alpha_{post} = 0.86$, $M_{post} = 4.24$, $SD_{post} = 0.95$) being exposed to why Casey had voted the way they did.

Persuasiveness of story

We asked six questions to measure Casey's rationale for voting for the other party persuasiveness on a 7-point scale (1, *Not at all*, to 7, *Extremely*). Four items were from a prior study that measured political persuasion (Schroeder et al., 2017; e.g., "How much do you think your beliefs changed as a result of Casey?"), and we added two additional questions assessing the impression of the participant overall (e.g., "How positive is your overall impression of Casey?"). These items were averaged to create an index of persuasiveness ($\alpha = 0.85$ for Republican stories, $\alpha = 0.83$ for Democratic stories; $\alpha = 0.85$, $M = 3.49$, $SD = 1.33$).

Analytic plan

We had two dependent variables of interest: interest in understanding perspective (provided prior to reading) and the dehumanization of their partner (measured both before and after reading the story). The first dependent variable—interest in understanding perspective—was analyzed using linear regression predicting interest by the experimental condition of whether Casey disrupted their highest-ranked meta-perception.

The second dependent variable—dehumanization—was analyzed using a linear mixed model, with each participant acting as their own random effect, where we predict dehumanization by the effect of the experimental condition of disruption, the experimental condition of how Casey's story was presented (voice or text), the time (before and after the story presentation), and their interactions. Models that account for the within-subjects factor of time can result in decimal-point degrees of freedom (Gaylor and Hopper, 1969). We did not have any leading hypotheses or reasons to expect a three-way interaction, so we do not model it in our models, nor do we include the interaction of time and method. Since the method-of-story independent variable does not

TABLE 2 Correlation matrix for study 1.

	Dehumanization	Persuasiveness	Age
Interest	-.39***	.39***	.08
Dehumanization	–	.57***	-.02
Persuasiveness	–	–	-.20***

Probability values are adjusted for multiple comparisons. *** $p < .001$.

differ at the pre-level (because participants were not yet exposed to what method they may receive), there would be no meaningful interaction to examine.

Our third measure, the persuasiveness of the story, acts as a manipulation check of the stimuli.

Across analyses, results do not change in significance if controlling for gender, political party, or interparty dehumanization. Thus, for simplicity in results, we report analyses without controls.² All variables are standardized to have a mean of zero and a standard deviation of 1 prior to being entered into their linear models, but the mixed-level models do not include the standardization of the variables due to concerns over standardizing on repeated measure designs (Moeller, 2015). Instead, to calculate effect sizes of the mixed-effect models, we take the approach outlined by Westfall et al. (2014), whereby the effect size (d) can be calculated as the difference between the means (the estimate for the fixed effect) divided by the square root of the sum of variances of random effects (Brybaert and Stevens, 2018).

A preliminary inspection of the data showed that several of the participants' text responses were merely copied and pasted from Casey's story. To filter these out, we calculated what percentage of the words they used did not appear in Casey's story and refer to this as a uniqueness score. Almost all participants had some overlap in their text with Casey's (e.g., using the words *I* or *and*). Responses from participants with a uniqueness of <50% were examined by hand. Those who had clearly copied Casey's story word for word with only a little embellishment were removed from the sample. In Study 1, there were six scores below 50%. Of these, four had a uniqueness score of 0%, one had a uniqueness score of 21.7%, and one had a uniqueness score of 100% and was entirely non-sense words. We remove these observations from our analyses,³ leaving the final N at 367.

Results

Table 2 shows the correlation matrix of all variables of interest with significance controlled for multiple comparisons.

Interest in understanding perspective

We ran a linear model predicting participants' interest in learning more about Casey's reasoning [$F_{(1,355)} = 4.996$, $p = .026$,

² For full analyses including the controls, see our OSF.

³ Significance and effect sizes do not meaningfully change with or without excluding these participants. See our OSF for their responses and analyses with the inclusion of the participants.

$r_{adj}^2 = .011$]. Having one's meta-perception disrupted increased one's interest [$\beta = 0.24$, $SE = 0.11$, $t_{(355)} = 2.24$, $p = .026$].

Persuasiveness

A linear model predicting participant's persuasiveness ratings of the story by the interaction of story presentation condition and meta-perception disruption condition showed no significant results [$F_{(3,355)} = 1.744$, $p = .158$, $r_{adj}^2 = .006$]. Stories were not judged differently in their persuasiveness regardless of the meta-perception disruption condition [$\beta = 0.21$, $SE = 0.14$, $t_{(355)} = 1.47$, $p = .143$]; if they heard the rationale compared to reading it [$\beta = -0.11$, $SE = 0.15$, $t_{(355)} = -0.76$, $p = .449$]; or the interaction of disruption by voice [$\beta = 0.01$, $SE = 0.21$, $t_{(355)} = 0.04$, $p = .966$].

Dehumanization

Participants were not more likely to attribute more humanity to stories read by a voice compared to reading it themselves [$b = 0.13$, $SE = 0.12$, $t_{(357)} = 1.09$, $p = .276$, $d = 0.12$]. However, there was a significant effect of both disruption condition and time. Participants dehumanized Casey less if they were exposed to the disrupted meta-perception condition [$b = -0.35$, $SE = 0.12$, $t_{(433.3)} = -2.81$, $p = .005$, $d = 0.31$], and after having read Casey's reasoning [$b = -0.34$, $SE = 0.05$, $t_{(359)} = -6.38$, $p < .001$, $d = 0.3$]. There was no significant moderation of disruption condition by time [$b = 0.03$, $SE = 0.08$, $t_{(359)} = 0.41$, $p = .684$, $d = 0.03$], or method of rationale transmission [$b = 0.04$, $SE = 0.17$, $t_{(357)} = 0.21$, $p = .832$, $d = 0.04$] (see [Figure 1](#)).

Discussion

In this study, we exposed individuals to someone who either does or does not disrupt their highest-held meta-perception. When exposed to an individual who disrupted their meta-perception, participants were more interested in learning more about their partner and dehumanized the individual less. We also found an equal effect of time, whereby learning about the personal experience of why an individual had voted for the other side also decreased the dehumanization of the partner. These effects are not due to the story itself, as the story was rated equally as persuasive regardless of meta-perception conditions.

However, we failed to replicate the effect of increased humanization when hearing the voice of a political opponent, contrary to prior work ([Schroeder et al., 2017](#)). Participants exposed to hearing Casey's voice on why they voted for the opposing party were equally likely to humanize them compared to reading why Casey had voted for the opposing party. As participants had to pass attention checks to move onto the post-dehumanization measure, it also seems unlikely that those in one condition paid more or less attention to Casey's reasoning than those in the other condition.

This study provides beginning evidence of a new, individual-focused meta-perception correction intervention by exposing participants to someone who violates their highest-held meta-perception. The design of the study aimed to target not simply one's assumption of a single question (e.g., "How much do you believe the other side dehumanizes you?") but also an assumption about

the opposing group that the participant believes the most. Our design is novel by virtue of considering the strength of the belief being intervened on and directing the perception to more plausible, lived experience meta-perceptions.

While we showed that participants were interested in learning more and dehumanized Casey less, we did not test to see if our intervention could result in wanting to have future interactions with Casey—a critical next step in reducing political polarization. We test this directly in Study 2.

Study 2

Since Study 1 showed a non-significant result of the manipulation of voice or text, we opted to simplify the design and remove that manipulation. We also seek to examine whether our intervention in correcting one's meta-perception can result in an increased willingness to engage in future interactions with their partner.

Participants and procedure

Through Amazon Mechanical Turk, 200 participants were recruited. Due to discrepancies in their text responses to Casey,⁴ 10 participants were removed, following the procedure outlined in Study 1. Of these, six had a uniqueness score of 0%, while the other four had uniqueness scores <19%. A manual check of the responses also caught one response that was entirely copied and pasted from the Wikipedia page about Senator Bob Casey. As such, we remove these 11 observations from our analyses.

Therefore, 189 participants were analyzed (95 female, 94 male; $M_{age} = 44.04$, $SD_{age} = 12.77$). Of these 189, 127 identified as Democrat, and 62 identified as Republican. Participants identified were 0.53% African, 6.35% Black, 85.19% White, 1.06% Hispanic, 2.65% Native American, 1.06% South Asian, and 3.17% Southeast Asian.

The procedure repeats Study 1, with the only changes being no manipulation of the method of story (all stories are presented as text) and the inclusion of a measure that may better tap into interest in engaging with Casey in the future, presented both before and after reading the rationale of Casey's vote choice.

Measures

Meta-perceptions

The same list of potential meta-perceptions was provided as in Study 1.

Interest in understanding perspective

The scale remains the same as what was used in Study 1 ($\alpha = .89$, $M = 4.35$, $SD = 1.65$).

⁴ Significance and effect sizes do not meaningfully change with or without excluding these participants; see our OSF for full analyses.

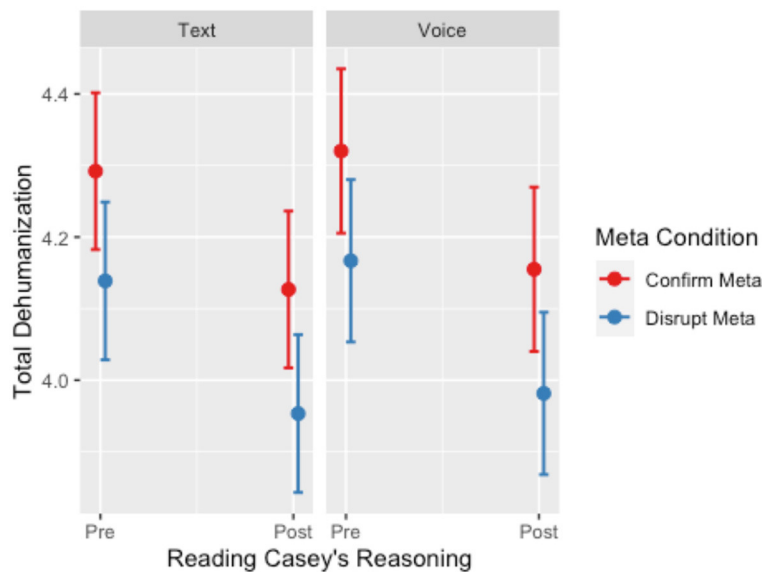


FIGURE 1
Predicted total dehumanization by condition and time in Study 1. Error bars represent the 95% confidence intervals.

Interest in future interactions

We created five items using a 7-point scale from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*) that measured the participants' future interaction intentions both before ($\alpha_{pre} = 0.89$, $M_{pre} = 4.1$, $SD_{pre} = 1.51$) and after ($\alpha_{post} = 0.91$, $M_{post} = 4.03$, $SD_{post} = 1.75$), reading Casey's rationale for voting for the opposing political party. An example of an item in this scale would be "If Casey sought contact with me, I would not respond" (reverse-coded).

Subtle dehumanization of Casey

The same scale that was used in Study 1 was used, both prior and after reading Casey's rationale (Bastian and Haslam, 2010) ($\alpha_{pre} = 0.89$, $M_{pre} = 4.01$, $SD_{pre} = 0.97$; $\alpha_{post} = 0.93$, $M_{post} = 4.46$, $SD_{post} = 1.13$).

Persuasiveness

The same scale used in Study 1 was used to measure the persuasiveness of the story ($\alpha = 0.89$ for Republican stories, $\alpha = 0.85$ for Democratic stories; $\alpha = 0.87$, $M = 3.8$, $SD = 1.42$).

Results

Table 3 reports the correlation matrix for the variables of interest. Across analyses, results do not change in significance if controlling for gender, political party, or interparty dehumanization. Thus, for simplicity, in the results, we report

analyses without controls.⁵ Models that account for the within-subjects factor of time are linear mixed models using restricted maximum likelihood estimation of the parameters, treating each individual participant as their own random effect.

Persuasiveness

A linear model predicting participants' persuasiveness ratings of the rationale by the meta-perception condition showed no significant results [$F_{(1,187)} = 0.961$, $p = 0.328$, $r_{adj}^2 = 0$]. Stories were not judged differently in their persuasiveness regardless of the disruption condition of the participant [$\beta = 0.14$, $SE = 0.15$, $t_{(187)} = 0.98$, $p = .328$].

Interest in understanding perspective

We ran a linear model predicting participant's interest in understanding Casey's perspective by whether or not Casey disrupted the meta-perception [$F_{(1,187)} = 6.439$, $p = .012$, $r_{adj}^2 = .028$]. Having one's meta-perception disrupted increased one's interest in understanding Casey's perspective [$\beta = 0.36$, $SE = 0.14$, $t_{(187)} = 2.54$, $p = .012$].

Dehumanization

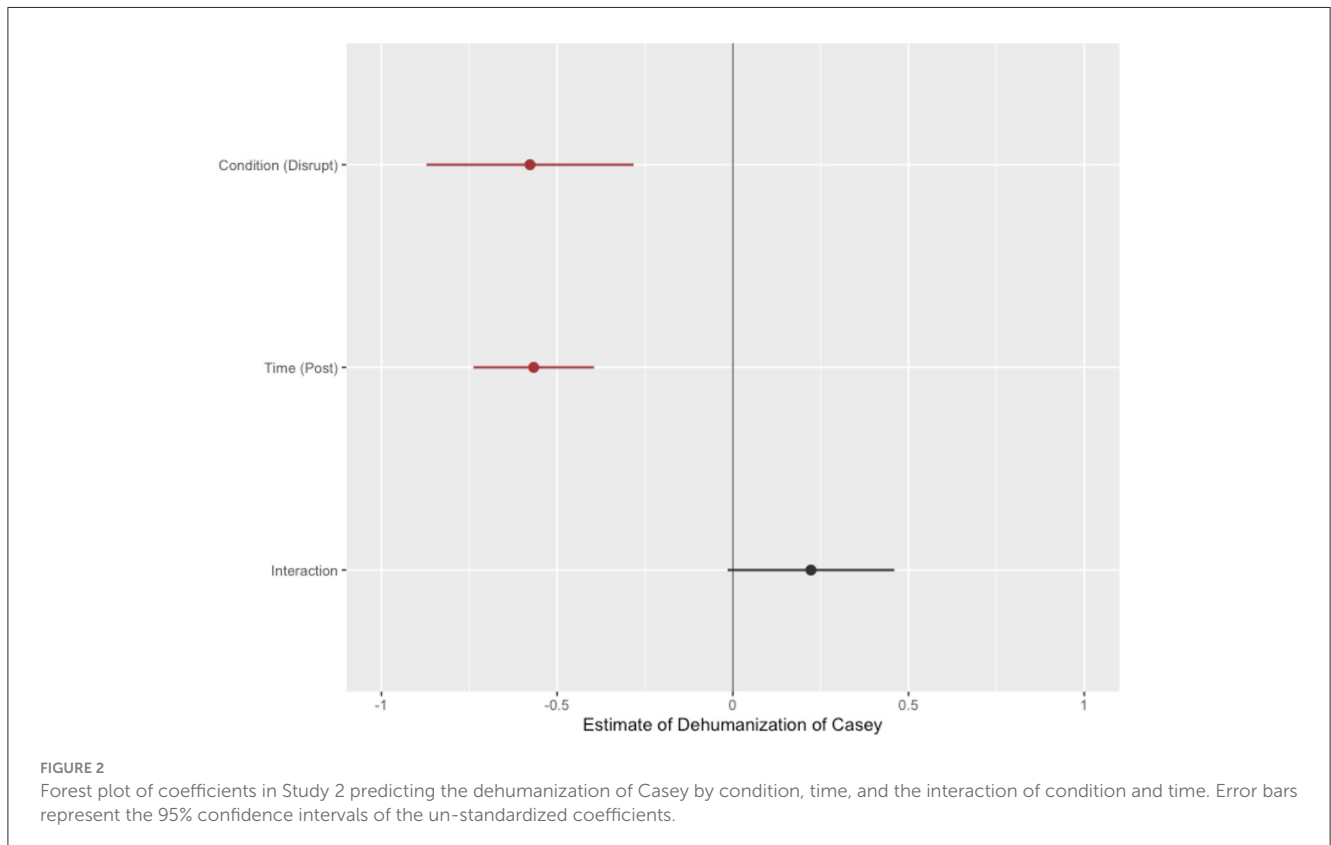
Participants' dehumanization of Casey was analyzed using a linear mixed model, with the betas and their confidence intervals plotted in Figure 2. Participants dehumanized Casey less after having been exposed to Casey's story [$b = -0.57$, $SE = 0.09$, $t_{(187)} = -6.49$, $p < .001$, $d = 0.59$], and if Casey disrupted their meta-perception [$b = -0.58$, $SE = 0.15$, $t_{(256.6)} = -3.85$, $p < .001$, $d =$

⁵ See our OSF for analyses including controls.

TABLE 3 Correlation matrix for study 2.

	Dehumanization	Persuasiveness	Future interaction	Age
Interest in Understanding	-.51***	.44***	.68***	-.08
Dehumanization	-	-.63***	-.63***	.07
Persuasiveness	-	-	.72***	-.19**
Future Interaction	-	-	-	-.14*

p-values are adjusted for multiple comparisons. **p* < .05. ***p* < .01. ****p* < .001.



0.6]. The interaction of Condition and Time was non-significant [*b* = 0.22, *SE* = 0.12, *t*₍₁₈₇₎ = 1.84, *p* = .067, *d* = 0.23].

Interest in future interactions

Participants' interest in interacting with Casey was analyzed using a linear mixed model. The resulting regression showed that neither disruption condition [*b* = 0.32, *SE* = 0.24, *t*_(219.6) = 1.36, *p* = .174, *d* = 0.52; time, *b* = 0.01, *SE* = 0.1, *t*_(184.2) = 0.06, *p* = .952, *d* = 0.02]; nor the interaction of disruption condition and time was significant [*b* = -0.13, *SE* = 0.14, *t*_(184.3) = -0.93, *p* = .352, *d* = 0.21].

In an exploratory analysis, we analyzed whether a participant's interest in reading Casey's rationale moderated the disrupted meta-perception condition to lead to interest in future interactions. The resulting regression showed that neither disruption condition [*b* = 0.74, *SE* = 0.46, *t*_(192.5) = 1.6, *p* = .111, *d* = 0.88; time, *b* = 0.01, *SE* = 0.1, *t*_(184.4) = 0.09, *p* = .929, *d* = 0.01]; nor the interaction of disruption condition and time was significant [*b* = -0.13, *SE* =

0.14, *t*_(184.6) = -0.95, *p* = .343, *d* = 0.15]. However, a willingness to learn about Casey's perspective was predictive of an interest in future interactions [*b* = 0.78, *SE* = 0.07, *t*₍₁₈₄₎ = 11.09, *p* < .001, *d* = 0.93], and this was not moderated by condition [*b* = -0.19, *SE* = 0.1, *t*_(184.4) = -1.93, *p* = .056, *d* = 0.23].

Discussion

In this study, we simplify the methods of Study 1 and replicate whether being exposed to an individual who disrupts one's meta-perception reduces dehumanization, increases interest in learning more, and increases willingness for future interactions. Of these goals, we successfully replicate the first two. Participants exposed to an individual who disrupts their meta-perception are dehumanized less compared to individuals who have their meta-perceptions confirmed, and participants are more likely to show interest in learning more about a political opponent if they disrupt our meta-perception.

However, our disruption intervention and learning about the partner's rationale failed to increase interest in future interactions. This null finding is itself in contraction with prior research. For example, in one study, informing individuals that both sides overestimate the meta-dehumanization of the opposing party led participants to report less desire for social distance, defined as the reported comfort for their doctor, child's teacher, and child's friend being a member of the opposing party through a decreased reported dehumanization of the other side and this result persisted a week later (Landry et al., 2023). In a larger multi-site study, correcting misperceptions significantly reduced desired social distance (Voelkel et al., 2023, Figure S8.4). Our operationalization of intergroup contact was more overt and direct than the indirect social contact measured in these social distance measures, especially since there is a reluctance to engage in intergroup contact with hated groups (Ron et al., 2017). A more direct and overt attempt to measure intergroup contact interest may not have captured the smaller movements toward reducing social distance compared to the aforementioned studies. Individuals are also poor at estimating the negative affect they would experience being exposed to a member of the opposing side (Dorison et al., 2019), and it could be that individuals were heightened in their emotional anxiety and chose to avoid further contact.

Exploratory analysis on message back to Casey

Studies 1 and 2 showed that disrupting one's meta-perceptions leads to decreased dehumanization and increased interest in learning more about your political opponent through classic quantitative methodology. However, developments in political science suggest that we can utilize text analysis to predict the characteristics of speakers such as their political ideology (Diermeier et al., 2012). Therefore, in an exploratory analysis, we pooled the text responses from Studies 1 and 2 ($N = 556$) in order to examine if we could use the sentiment of the written responses to predict whether Casey had confirmed or disrupted the participant's highest-ranked meta-perception.

We calculated emotion scores for each response using the R library *sentimentr*. These scores are calculated using a valence dictionary for each of eight emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) and the negation of each. A response's score for each emotion is the number of words in the response that appear in the anger dictionary divided by the total number of words in the response and is thus a number from 0 to 1. We used the default NRC hashtag Emotion Lexicon, which was trained on tweets tagged with the relevant emotion. This approach is similar to how the LIWC software calculates its own scores (Pennebaker et al., 2003).

To see if different kinds of emotion are associated with responding to different kinds of political ideology, we then fit a logistic regression model predicting whether Casey strongly agreed or disagreed. Only a single sentiment value had a $p < 0.05$ threshold (disgust negated; $Z = 2.126$; $p < 0.05$). But with 16 variables tested, one predictor with such a value is unsurprising if we assume no statistical significance. It is possible that such an analysis could be improved by building a dictionary specifically for this domain, but that would require a much larger corpus than we have.

Instead of using a pre-built dictionary, we considered whether each individual word in the participants' responses indicated whether Casey strongly agreed or disagreed. To do this, we first filtered out words that were used by respondents fewer than 10 times. We then considered what percentage of respondents in each category used each word. To avoid infinite ratios later, the minimum number of times a word could have in each group was manually set to 1. This approach is notably different than LIWC-based analyses as we use the words themselves—not an assigned score based on a percentage of words given a prebuilt dictionary—to predict the condition.

Figure 3 presents the log ratios of the frequency of usage for the 10 words that are most strongly associated with responses, grouped by Casey's agreement. A log ratio of 0 means that the frequencies were exactly equal for each group, a value of +1 (−1) means that participants responding to a Casey who strongly agreed (disagreed) used the word approximately 2.7 times as frequently as those responding to a Casey who strongly disagreed (agreed).

Participants responding to a Casey who strongly agreed used words disengaging from and putting a distance between and Casey. Three of the top five words are *mine*, *whatever*, and *society* (the word *united* is only used to reference the United States, and *guns* is used to reference topics that were specifically prompted, and so are not discussed here.) The word *mine* appears repeatedly in an agree-to-disagree context, as is seen in “Casey can have opinions and I will have mine” and “Your perspective is yours and not mine to judge.” *Whatever* is also seen in a dismissive context in the responses “Everyone is free to make whatever choices they want” and “such as the way that he (Trump) said whatever he felt like.” The words *beliefs* (“Casey is entitled to his beliefs and obviously, they hold much value to him”) and *judge* (quoted earlier) show similar usage.

When Casey disagreed, the clear narrative is less clear. The top term by far is *thousands*, a term used both by Democrats and Republicans in reference to big problems like “thousands of illegals” and “pandemic taking thousands of lives.” Here we see participants' frustration that, while Casey disagrees with the participant's meta-perceptions about Casey's party, they still have chosen to back a candidate the participant strongly dislikes. By comparison, *valid* shows the opposite result, with participants stating that “how you made your choices had valid reasons.”

Discussion

In this article, we tested whether we could decrease intergroup hostilities by focusing on correcting meta-perceptions—inaccurate assumptions of what an out-group thinks of our in-group. In two studies, we found that disrupting meta-perceptions decreased dehumanization and increased the level of interest in learning more about one's partner but failed to increase reported willingness to engage in future intergroup contact. We then explored the language use between our conditions to begin a deeper understanding of how these interactions are occurring at a message-sending level.

We advance current research in three incremental steps. First, our design does not simply inform participants of their own misperceptions, as is the operating standard (Harper and Baguley, 2022; Mernyk et al., 2022). Instead, we provide participants an

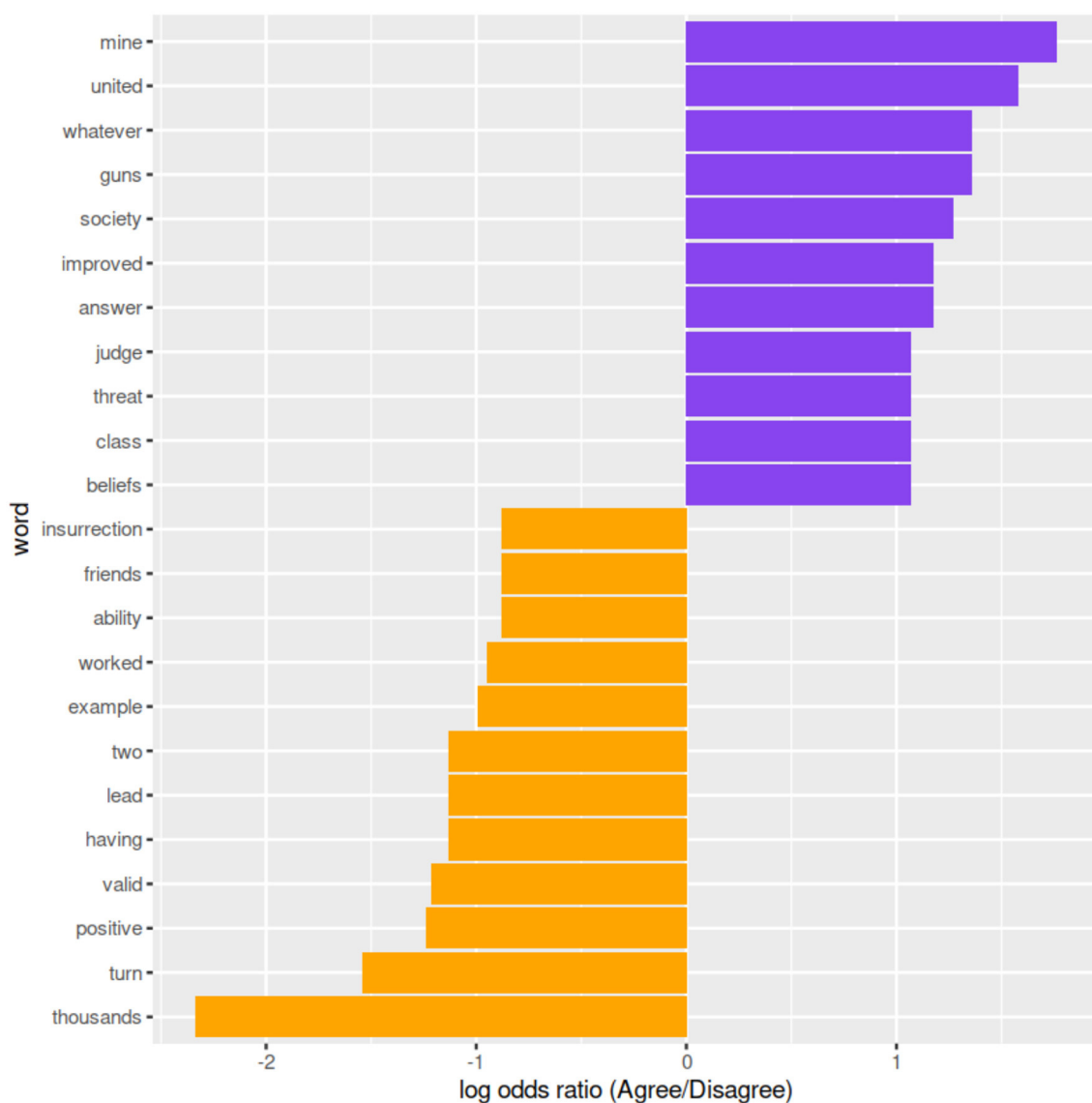


FIGURE 3 Log odds ratio of word use across Studies 1 and 2 in messages sent back to Casey, predicting their condition.

opportunity to select from a list of possible meta-perceptions that they may hold and invite them to meet someone who confirms or disrupts that belief. While a more direct solution to misperception correction, it does leave open the possibility that participants held none of those beliefs in particular. We were unable to confirm the strength at which participants held these beliefs (only that it was the highest ranked from the provided list), so future work could elaborate on this by collecting and controlling for this. This would be important because meta-perception interventions have shown to be strongest for highly partisan, highly inaccurate individuals (Mernyk et al., 2022; Pasek et al., 2022).

Second, the correction itself is not a response on a scale but, instead, plausible facts one may believe or encounter through the media. We find such a manipulation to hold more external validity than estimations on the Ascent of Man scale. Finally, we move from a group-level correction

toward an individual-level correction by presenting participants with an exemplar who disproves the rule. It also would be fruitful for future work to test this manipulation in person. Instead of engaging with Casey through a Qualtrics survey, confederates could memorize the script and meet the participant face-to-face. For example, Kubin et al. (2021) used in-person interactions on political topics to show the power of personal experiences instead of facts in increasing respect for a political adversary.

Study 2 failed to show a downstream effect of increasing future intergroup contact. As we outline in the discussion of Study 2, this null effect may be due to measurement differences. Prior interventions that have tried to measure intergroup constructs beyond dehumanization and stereotyping have focused on social distance (Landry et al., 2023; Voelkel et al., 2023), anti-democratic norms (Moore-Berg et al., 2020), out-group hostility (Landry

et al., 2022), and belief in out-group obstructionism as a goal (Lees and Cikara, 2020). While participants were interested in learning why the disrupted condition voted for the other side more than the confirmed side and interest was predictive of future interaction intention, there was no moderating effect of meta-perception correction on interaction intention by interest. It could simply be that the measure of intergroup contact intent was too severe, and smaller incremental steps would need to be measured first. As Landry et al. (2023) note, “correcting exaggerated meta-dehumanization produced far stronger reductions in reciprocal dehumanization than these downstream outcomes” (p. 414), reflecting the current difficulty in translating reductions in dehumanization to further important intergroup constructs.

A large limitation in this research is an inability to know whether or not the meta-perceptions of the participants were truly changed—that is, whether participants acknowledged their meta-perceptions were wrong or if they discounted the individual as simply an outlier that represents the exception to the rule and kept their operating meta-perception intact through subtyping the individual away from the targeted out-group (Richards and Hewstone, 2001; Bott and Murphy, 2007). We did not measure whether they still held that meta-perception after meeting their partner, so in this article, we cannot speak to whether the perceptions were fully corrected. This limitation of subtyping could have also been addressed if we had collected the dehumanization of the political party after either learning that Casey disrupted the meta-perception or after learning of Casey’s rationale for voting for the opposing party. However, because we only collected dehumanization of the opposing political parties at the beginning of the study prior to any exposure, our measure of interparty dehumanization cannot address this concern. Future research could measure interparty dehumanization after meta-perception disruption to investigate if dehumanization of the party decreased as a whole.

However, we are tentatively optimistic about the chances of updating one’s meta-perceptions. A recent meta-analysis showed that exemplars can update stereotypes, especially if said exemplars are moderately different from the stereotype in comparison to being extremely different (McIntyre et al., 2016). The chosen rationale reported by the partner was designed with moderation in mind, having been piloted to be the most persuasive of possible rationales, and we found no differences in how individuals perceived these rationales. This does not explicitly measure the updating of one’s meta-perceptions (and a pre/post measure of one’s top-rated meta-perceptions may more directly target this question), but we are hopeful that the rationale was not viewed as too extreme. Future research would need to measure and clarify this question before we can fully conclude this research as another piece of evidence on how atypical interventions can reduce intergroup conflict and bias (Dasgupta and Greenwald, 2001; Dasgupta and Asgari, 2004; Yetkili et al., 2018; Kurdi et al., 2023; Maunder and White, 2023).

We also found no significant main or moderating effect of the method of transmission of Casey’s rationale being presented as text or voice. Regardless of whether participants read or listened to Casey’s rationale for voting for the other side, it had no impact on how much humanness the participants reported. Participants had to pass an attention check on why Casey voted the way they

did, suggesting that it was not due to the participants not paying attention. The prior work that we were drawing from was also done on online samples (Schroeder et al., 2017), so more research needs to be done to sort out the discrepancy between the results, especially in light of their second study having their participants report why they supported particular candidates being extremely similar to our own.

We also used text analysis as an exploratory framework to explore how our intervention qualitatively changed participant responses to Casey. We chose an exploratory, descriptive analysis due to the high risk of false positives due to the vast number of words that could have been used. The descriptive analysis provided interesting patterns that are worthy of further exploration. In particular, the differences in language use suggest that our understanding of the social norms of others molds current and future conversations. Importantly, our manipulation did not avoid tense topics—words like *insurrection* were still more likely to occur within the disrupted condition. However, we think incorporating text analysis has a bright future in psychological research (see Kubin et al., 2021, Study 3, for text analysis of YouTube comments), given appropriate precautions of false positives and multiple corrections, and provide interesting avenues for more descriptive data analysis for future theory building (Power et al., 2018; Hofmann and Grigoryan, 2023; Bonetto et al., 2024).

Overall, our work provides a new direction in resolving misperceptions of others by creating opportunities to meet an individual who disrupts one’s previously held beliefs. Our results show that we can achieve decreased dehumanization and increased interest in learning about the partner’s rationale. However, our intervention was unable to get participants to report wanting to meet in future interactions. Still, we hold this as a good attempt to clarify the true minds of the other and work to build more bridges between groups.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://doi.org/10.17605/OSF.IO/XQRTN>.

Ethics statement

The studies involving humans were approved by the Washington and Jefferson Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources,

Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. GH: Formal analysis, Methodology, Visualization, Writing – review & editing. DS: Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing – review & editing. RG: Conceptualization, Data curation, Investigation, Methodology, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work received support from the Institute for Humane Studies under grant no. IHS017893.

References

- Ahler, D. J. (2014). Self-fulfilling misperceptions of public polarization. *The J. Polit.* 76, 607–620. doi: 10.1017/S0022381614000085
- Ahler, D. J., and Sood, G. (2018). The parties in our heads: misperceptions about party composition and their consequences. *The J. Politics* 80, 964–981. doi: 10.1086/697253
- Alexander, M. G., Brewer, M. B., and Livingston, R. W. (2005). Putting stereotype content in context: image theory and interethnic stereotypes. *Pers. Soc. Psychol. Bull.* 31, 781–794. doi: 10.1177/0146167204271550
- Bastian, B., and Haslam, N. (2010). Excluded from humanity: the dehumanizing effects of social ostracism. *J. Exp. Soc. Psychol.* 46, 107–113. doi: 10.1016/j.jesp.2009.06.022
- Baumeister, R. F., Vohs, K. D., and Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Pers. Psychol. Sci.* 2, 396–403. doi: 10.1111/j.1745-6916.2007.00051.x
- Bonetto, E., Guiller, T., and Adam-Troian, J. (2024). A lost idea in psychology: Observation as starting point for the scientific investigation of human behavior. *PsyArXiv*. doi: 10.31234/osf.io/9yk3n
- Bott, L., and Murphy, G. L. (2007). Subtyping as a knowledge preservation strategy in category learning. *Mem. Cognit.* 35, 432–443. doi: 10.3758/BF03193283
- Broockman, D. E., Kalla, J. L., and Westwood, S. J. (2023). Does affective polarization undermine democratic norms or accountability? Maybe not. *Am. J. Polit. Sci.* 67, 808–828. doi: 10.1111/ajps.12719
- Bruneau, E., Hameiri, B., Moore-Berg, S. L., and Kteily, N. S. (2021). Intergroup contact reduces dehumanization and meta-dehumanization: cross-sectional, longitudinal, and quasi-experimental evidence from 16 samples in five countries. *Pers. Soc. Psychol. Bull.* 47, 906–920. doi: 10.1177/0146167220949004
- Brybaert, M., and Stevens, M. (2018). Power analysis and effect size in mixed effects models: a tutorial. *J. Cognit.* 1:9. doi: 10.5334/joc.10
- Connaughton, A. (2021). *Americans See Stronger Societal Conflicts Than People in Other Advanced Economies*. Pew Research Center. Available online at: <https://www.pewresearch.org/short-reads/2021/10/13/americans-see-stronger-societal-conflicts-than-people-in-other-advanced-economies/> (accessed October 13, 2021).
- Dasgupta, N., and Asgari, S. (2004). Seeing is believing: exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *J. Exp. Soc. Psychol.* 40, 642–658. doi: 10.1016/j.jesp.2004.02.003
- Dasgupta, N., and Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *J. Pers. Soc. Psychol.* 81, 800–814. doi: 10.1037/0022-3514.81.5.800
- Diermeier, D., Godbout, J. F., Yu, B., and Kaufmann, S. (2012). Language and ideology in Congress. *Br. J. Polit. Sci.* 42, 31–55. doi: 10.1017/S0007123411000160
- Dorison, C. A., Minson, J. A., and Rogers, T. (2019). Selective exposure partly relies on faulty affective forecasts. *Cognition* 188, 98–107. doi: 10.1016/j.cognition.2019.02.010
- Eisenkraft, N., Effenbein, H. A., and Kopelman, S. (2017). We know who likes us, but not who competes against us: dyadic meta-accuracy among work colleagues. *Psychol. Sci.* 28, 233–241. doi: 10.1177/0956797616679440
- Finchilescu, G. (2010). Intergroup anxiety in interracial interaction: the role of prejudice and metastereotypes. *J. Soc. Issues* 66, 334–351. doi: 10.1111/j.1540-4560.2010.01648.x
- Frey, F. E., and Tropp, L. R. (2006). Being seen as individuals versus as group members: Extending research on metaperception to intergroup contexts. *Pers. Soc. Psychol. Rev.* 10, 265–280. doi: 10.1207/s15327957pspr1003_5
- Gaylor, D. W., and Hopper, F. N. (1969). Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics* 11, 691–705. doi: 10.1080/00401706.1969.10490732
- Harper, C. A., and Baguley, T. (2022). “You are fake news”: Ideological (a)symmetries in perceptions of media legitimacy. *PsyArXiv*. doi: 10.31234/osf.io/ym6t5
- Hodson, G., Crisp, R. J., Meleady, R., and Earle, M. (2018). Intergroup contact as an agent of cognitive liberalization. *Pers. Psychol. Sci.* 13, 523–548. doi: 10.1177/1745691617752324
- Hofmann, W., and Grigoryan, L. (2023). The social psychology of everyday life. *Adv. Exp. Soc. Psychol.* 68, 77–137. doi: 10.1016/bs.aesp.2023.06.001
- Kteily, N. S., and Bruneau, E. (2017). Backlash: the politics and real-world consequences of minority group dehumanization. *Pers. Soc. Psychol. Bull.* 43, 87–104. doi: 10.1177/0146167216675334
- Kteily, N. S., Bruneau, E., Waytz, A., and Cotterill, S. (2015). The ascent of man: theoretical and empirical evidence for blatant dehumanization. *J. Pers. Soc. Psychol.* 109, 901–931. doi: 10.1037/pspp0000048
- Kteily, N. S., Hodson, G., and Bruneau, E. (2016). They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. *J. Pers. Soc. Psychol.* 110, 343–370. doi: 10.1037/pspa0000044
- Kubin, E., Puryear, C., Schein, C., and Gray, K. (2021). Personal experiences bridge moral and political divides better than facts. *Proc. Nat. Acad. Sci.* 118:e2008389118. doi: 10.1073/pnas.2008389118
- Kurdi, B., Sanchez, A., Dasgupta, N., and Banaji, M. R. (2023). (When) do counterattitudinal exemplars shift implicit racial evaluations? Replications and extensions of Dasgupta and Greenwald (2001). *J. Pers. Soc. Psychol.* doi: 10.1037/pspa0000370 [Epub ahead of print].
- Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., et al. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *J. Exp. Psychol. General* 143, 1765–1785. doi: 10.1037/a0036260
- Landry, A. P., Ihm, E., and Schooler, J. W. (2022). Hated but still human: metadehumanization leads to greater hostility than metaprejudice. *Group Proc. Interg. Relations* 25, 315–334. doi: 10.1177/1368430220979035
- Landry, A. P., Schooler, J. W., Ihm, E., and Kwit, S. (2021). Metadehumanization erodes democratic norms during the 2020 presidential election. *Anal. Soc. Issues Pub. Policy* 21, 51–63. doi: 10.1111/asap.12253
- Landry, A. P., Schooler, J. W., Willer, R., and Seli, P. (2023). Reducing explicit blatant dehumanization by correcting exaggerated meta-perceptions. *Soc. Psychol. Pers. Sci.* 14, 407–418. doi: 10.1177/19485506221099146
- Lees, J. M., and Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nat. Hum. Behav.* 4, 279–286. doi: 10.1038/s41562-019-0766-4
- Mauder, R. D., and White, F. A. (2023). Exemplar typicality in interventions to reduce public stigma against people with mental illness. *J. Appl. Soc. Psychol.* 53, 819–834. doi: 10.1111/jasp.12970

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- McIntyre, K., Paolini, S., and Hewstone, M. (2016). Changing people's views of outgroups through individual-to-group generalisation: meta-analytic reviews and theoretical considerations. *Eur. Rev. Soc. Psychol.* 27, 63–115. doi: 10.1080/10463283.2016.1201893
- Mernyk, J. S., Pink, S. L., Druckman, J. N., and Willer, R. (2022). Correcting inaccurate metaperceptions reduces Americans' support for partisan violence. *Proc. Nat. Acad. Sci.* 119:e2116851119. doi: 10.1073/pnas.2116851119
- Moeller, J. (2015). A word on standardization in longitudinal studies: Don't. *Front. Psychol.* 6:1389. doi: 10.3389/fpsyg.2015.01389
- Moore-Berg, S. L., Ankori-Karlinsky, L. O., Hameiri, B., and Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proc. Nat. Acad. Sci.* 117, 14864–14872. doi: 10.1073/pnas.2001263117
- Mulderig, B., Carriere, K. R., and Wagoner, B. (2024). Memorials and political memory: a text analysis of online reviews. *PsyArXiv*. doi: 10.31234/osf.io/uxj8e
- Nomikos, W. G. (2022). Peacekeeping and the enforcement of intergroup cooperation: evidence from Mali. *The J. Polit.* 84, 194–208. doi: 10.1086/715246
- O'Brien, T. C., Leidner, B., and Tropp, L. R. (2018). Are they for us or against us? How intergroup metaperceptions shape foreign policy attitudes. *Group Proc. Interg. Relat.* 21, 941–961. doi: 10.1177/1368430216684645
- Overgaard, C. S. B., Dudo, A., Lease, M., Masullo, G. M., Stroud, N. J., Stroud, S. R., et al. (2021). "Building connective democracy," in *The Routledge Companion to Media Disinformation and Populism, 1st Edn*, eds H. Tumber and S. Waisbord (London: Routledge), 559–568.
- Pasek, M. H., Karlinsky, L. O. A., Levy-Vene, A., and Moore-Berg, S. (2022). Misperceptions about out-partisans' democratic values may erode democracy. *Sci. Rep.* 12:16284. doi: 10.1038/s41598-022-19616-4
- Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. *Ann. Rev. Psychol.* 54, 547–577. doi: 10.1146/annurev.psych.54.101601.145041
- Petsko, C. D., and Kteily, N. S. (2023). Political (meta-)dehumanization in mental representations: divergent emphases in the minds of liberals versus conservatives. *Pers. Soc. Psychol. Bull.* 12:01461672231180971. doi: 10.1177/01461672231180971
- Pew Research Center (2022). *As Partisan Hostility Grows, Signs of Frustration With the Two-Party System*. Pew Research Center. <https://www.pewresearch.org/politics/2022/08/09/as-partisan-hostility-grows-signs-of-frustration-with-the-two-party-system/> (accessed August 9, 2022).
- Power, S. A., Velez, G., Qadafi, A., and Tennant, J. (2018). The SAGE model of social psychological research. *Persp. Psychol. Sci.* 13, 359–372. doi: 10.1177/1745691617734863
- Prati, F., Vasiljevic, M., Crisp, R. J., and Rubini, M. (2015). Some extended psychological benefits of challenging social stereotypes: decreased dehumanization and a reduced reliance on heuristic thinking. *Group Proc. Interg. Relat.* 18, 801–816. doi: 10.1177/1368430214567762
- Richards, Z., and Hewstone, M. (2001). Subtyping and subgrouping: processes for the prevention and promotion of stereotype change. *Pers. Soc. Psychol. Rev.* 5, 52–73. doi: 10.1207/S15327957PSPR0501_4
- Ron, Y., Solomon, J., Halperin, E., and Saguy, T. (2017). Willingness to engage in intergroup contact: a multilevel approach. *Peace Conf. J. Peace Psychol.* 23, 210–218. doi: 10.1037/pac0000204
- Ruggeri, K., Večkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., et al. (2021). The general fault in our fault lines. *Nat. Hum. Behav.* 5, 1369–1380. doi: 10.1038/s41562-021-01092-x
- Saguy, T., and Kteily, N. S. (2011). Inside the opponent's head: perceived losses in group position predict accuracy in metaperceptions between groups. *Psychol. Sci.* 22, 951–958. doi: 10.1177/0956797611412388
- Schroeder, J., Kardas, M., and Epley, N. (2017). The humanizing voice: speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychol. Sci.* 28, 1745–1762. doi: 10.1177/095679761713798
- Shackelford, C. M., Pasek, M. H., Vishkin, A., and Ginges, J. (2024). Palestinians and Israelis believe the other's God encourages intergroup benevolence: a case of positive intergroup meta-perceptions. *J. Exp. Soc. Psychol.* 110:104551. doi: 10.1016/j.jesp.2023.104551
- Simon, A. F., and Wilder, D. (2022). Methods and measures in social and personality psychology: a comparison of JPSP publications in 1982 and 2016. *The J. Soc. Psychol.* 21, 1–17. doi: 10.1080/00224545.2022.2135088
- Stathi, S., Di Bernardo, D., Vezzali, G. A., Pendleton, L. S., and Tropp, L. R. (2020). Do they want contact with us? The role of intergroup contact meta-perceptions on positive contact and attitudes. *J. Commun. Appl. Soc. Psychol.* 30, 461–479. doi: 10.1002/casp.2452
- Tajfel, H., and Turner, J. C. (1979). "An integrative theory of intergroup conflict" in *The Social Psychology of Intergroup Relations*, eds S. Worchel and W. G. Austin (London: Brooks/Cole), 33–47.
- Taylor, K. (2007). Disgust is a factor in extreme prejudice. *Br. J. Soc. Psychol.* 46, 597–617. doi: 10.1348/014466606X156546
- Voelkel, J. G., Chu, J., Stagnaro, M. N., Mernyk, J. S., Redekopp, C., Pink, S. L., et al. (2023). Interventions reducing affective polarization do not improve anti-democratic attitudes. *Nat. Hum. Behav.* 7, 55–64. doi: 10.1038/s41562-022-01466-9
- Vorauer, J. D., Hunter, A. J., Main, K. J., and Roy, S. A. (2000). Meta-stereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *J. Pers. Soc. Psychol.* 78, 690–707. doi: 10.1037/0022-3514.78.4.690
- Vorauer, J. D., Main, K. J., and O'Connell, G. B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *J. Pers. Soc. Psychol.* 75, 917–937. doi: 10.1037/0022-3514.75.4.917
- Westfall, J., Kenny, D. A., and Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *J. Exp. Psychol. Gen.* 143, 2020–2045. doi: 10.1037/xge0000014
- Yetkili, O., Abrams, D., Travaglino, G. A., and Giner-Sorolla, R. (2018). Imagined contact with atypical outgroup members that are anti-normative within their group can reduce prejudice. *J. Exp. Soc. Psychol.* 76, 208–219. doi: 10.1016/j.jesp.2018.02.004
- Zingora, T., Vezzali, L., and Graf, S. (2021). Stereotypes in the face of reality: Intergroup contact inconsistent with group stereotypes changes attitudes more than stereotype-consistent contact. *Group Proc. Interg. Relations* 24, 1284–1305. doi: 10.1177/1368430220946816