

Exploring Different Psychological Processes in a Media Intervention That Reduces Dehumanization Towards Muslims

Roman A. Gallardo 

University of Pennsylvania
University of Chicago

Samantha L. Moore-Berg 

University of Pennsylvania
University of Utah

Boaz Hameiri

Tel Aviv University

Dehumanization continues to be prevalent today and predicts detrimental intergroup consequences. Thus, it is important to identify novel interventions that reduce dehumanization and explore the mechanism(s)—both established (e.g., empathy induction, intergroup contact) and relatively understudied (e.g., humor)—driving the effects. To address this issue, in Study 1 (N = 2,349), we conducted an “intervention tournament” and found that a video (i.e., “Mean Tweets”) of a relatable and diverse group of young Muslims ridiculing Islamophobic comments posted on an online video of a Muslim preschool burning down effectively reduces hostility towards Muslims. Specifically, the Mean Tweets intervention significantly reduced dehumanization of Muslims and, although the effects were weaker, anti-Muslim policy support. However, a follow-up study conducted 1 month later revealed that these effects subsided. Next, in a preregistered study (i.e., Study 2) (N = 677), we find that our intervention reliably reduced dehumanization of Muslims, but the reduction of anti-Muslim policy support was not replicated. While investigating our intervention’s effect on anti-Muslim policy support, we find that our intervention led participants to discount the intervention’s intended message, which could have short-circuited the intervention’s effectiveness on policy support. Considering these results, we discuss the potential psychological processes (e.g., humor, message discounting, tone) underlying our dehumanization-reducing intervention.

KEY WORDS: dehumanization, interventions, Islamophobia, intervention tournament

Dehumanization, or the denial of full humanness from individuals or groups, has shaped some of the darkest times in history and continues to be prevalent in intergroup relations today (Haslam, 2015; Haslam & Stratemeyer, 2016; Kteily & Bruneau, 2017; Steizinger, 2018). Moreover, research shows that dehumanization of a group predicts consequential aggressive

behavior towards the dehumanized group (e.g., support for torture, support of housing and educational discrimination, support for collective punishment) (Bruneau & Kteily, 2017; Jardina & Piston, 2021; Kteily et al., 2015). To address this issue, researchers have designed and tested interventions that can short-circuit dehumanization and prevent the associated detrimental outcomes (Bruneau et al., 2021; McDonald et al., 2017; Moore-Berg et al., 2022). Although researchers have identified various intervention approaches that reduce dehumanization (e.g., positive intergroup contact; Capozza, Falvo, Di Bernardo, et al., 2014, countering stereotypes; Prati et al., 2015), identifying effective interventions, based on established or new approaches, that reduce dehumanization is crucial as groups continue to be consistently dehumanized (Bruneau & Kteily, 2017; Pavetich & Stathi, 2021). In the current work, we sought to combat the dehumanization of Muslims using various approaches and explore the psychological processes underlying dehumanization-reducing interventions.

Research on Dehumanization and How to Combat It

Recent social psychology findings suggest that across a broad range of samples, there is a striking level of dehumanization towards groups, especially disadvantaged groups ranging from immigrants to racial and ethnic groups (e.g., Haslam & Stratemeyer, 2016; Kteily & Bruneau, 2017). For example, Kteily et al. (2015) found that Americans explicitly rated Muslims, Arabs, Chinese people, South Koreans, and Mexican immigrants as less “evolved” than Americans (see also Kteily & Bruneau, 2017). Additionally, research suggests that dehumanization is a unique predictor of a variety of severely consequential outcomes, including support for torture of Muslims and Arabs among non-Muslim and non-Arab Americans, support of housing and educational discrimination against the Roma among non-Roma Hungarians, and support for collective punishment of Palestinians among Jewish-Israelis (Kteily et al., 2016). More recently, research has found a strong link between dehumanization and support for aggressive discriminatory policies towards Muslims, such as banning Muslims from entering the United States and banning the opening of any new mosques (Bruneau et al., 2018). Thus, we argue that dehumanization’s prevalence and link with aggressive attitudes and behaviors merit designing and assessing interventions that have the potential to mitigate dehumanization.

There are various approaches to reduce dehumanization that have been scientifically tested (e.g., countering stereotypes; Prati et al., 2015; inducing empathy; McDonald et al., 2017; Moore-Berg et al., 2022; highlighting hypocrisy; Bruneau et al., 2018, 2020; Gallardo et al., 2022). One of the most tested approaches to reduce dehumanization is through vicarious or direct, virtual, and extended contact (for a recent review, see Paolini et al., 2021). For example, Stathi et al. (2017) found that positive direct contact between Greek and Turkish Cypriots was correlated with decreased dehumanization of Greek Cypriots among Turkish Cypriots. In another study, Bruneau et al. (2021) found that facilitated positive semester-long virtual contact between non-Muslim American and Muslim college students from North Africa, the Middle East, and Southeast Asia reduced Americans’ dehumanization of Muslims. In another line of work, Capozza, Falvo, Trifiletti, and Pagani (2014) found that heterosexual students who reported more extended or vicarious contact with gay male students and lesbian female students (i.e., knowing other heterosexual students that have positive relationships with gay male students and lesbian female students) reported less dehumanization of gay male students and lesbian female students.

Practitioner Approaches to Improving Intergroup Relations

Practitioners have also developed and utilized dehumanization-reducing interventions. Practitioners' approaches in many cases opt for compelling styles (e.g., relatable narratives, parasocial contact, humorous messages) that are usually based on the practitioner's specific expertise, experience, and intuition (Bar-Tal & Hameiri, 2020; IJzerman et al., 2020). However, practitioners' approaches oftentimes are not rigorously tested, and thus, it is unclear whether these interventions reduce dehumanization and if so, how (Bar-Tal & Hameiri, 2020; IJzerman et al., 2020). Additionally, many of these interventions are already being implemented in the field. Therefore, we argue that it is crucial to test and better understand the effects of these interventions in order to identify the most effective approaches to facilitate better intergroup relations (e.g., IJzerman et al., 2020; Saguy, 2018).

One common approach used by practitioners is reducing intergroup dehumanization with humor. While humor has been utilized and tested in various domains to shift attitudes (e.g., politics; e.g., Innocenti & Miller, 2016; marketing; e.g., Beard, 2005; business; e.g., Cooper, 2005; education; e.g., Fata et al., 2018), its potential to improve intergroup relations has been underexplored by researchers despite Gordon Allport referring to humor as the "missing ingredient" (1954, p. 437) in the prejudice puzzle (for a notable exception, see Nir & Halperin, 2019). Why might humor be an effective persuasion tool in interventions designed to improve intergroup relations? First, humor has the potential to increase participants' motivation to process a message (Zhang & Zinkhan, 2006). This is particularly important as interventions designed to improve intergroup relations usually focus on socially sensitive subjects (e.g., racism, sexism, homophobia) that can discourage participants' willingness to process the potentially aversive message(s) (e.g., Hay et al., 2015). Second, previous research has found that humorous messages can distinctly hold participants' attention over nonhumorous messages. For example, Blanc and Brigaud (2014) found that participants who observed humorous (as opposed to nonhumorous) health advertisements were more attentive to the message and found the message more convincing. Finally, humor reduces negative cognitions related to the message and increases positive affect, which can amplify message effectiveness (Cline & Kellaris, 2007).

Additionally, we argue that on top of the persuasive elements of humorous messages, group members can use humor to provide information about their ingroup, which can reduce dehumanization of that group. In particular, animalistic dehumanization is the act of denying someone uniquely human traits (Haslam, 2006), and since previous work on humor suggests that humor is a uniquely human trait that distinguishes humans from other species (Martin & Ford, 2018), it could be that humor might reduce animalistic dehumanization specifically. Thus, we argue that portraying a group capable of displaying a uniquely human trait, humor, may be especially effective in reducing the dehumanization of that group (see also Tapley, 2006).

Another commonly used intervention strategy by practitioners is reducing intergroup prejudice and dehumanization with parasocial contact, or portraying groups in a positive light through the media (Schiappa et al., 2005; Varma, 2020). Research on parasocial contact suggests that observing positive media portrayals of outgroup individuals or displaying outgroup individuals positively can improve intergroup relations (e.g., Dovidio et al., 2011), through affective (e.g., increased empathy, reduced affective prejudice; e.g., Massey et al., 2021) and cognitive (e.g., by providing counterstereotypical information, correcting misperceptions; e.g., Bruneau et al., 2022; Vezzali et al., 2022) psychological mechanisms. For example, in two separate

studies, Schiappa et al. (2005) found that exposure to television series with positive portrayals of gay males (i.e., *Six Feet Under* and *Queer Eye for the Straight Guy*) reduced prejudicial attitudes towards gay males. As another example, Varma (2020) found that by writing relatable narratives of marginalized communities, American journalists reduced dehumanization of these respective communities. This provides some evidence that interventions utilizing parasocial contact could reduce dehumanization towards stigmatized groups. Finally, given that dehumanization is strongly associated with support for discriminatory policies, we argue that interventions that humanize a stigmatized group may in turn reduce support for discriminatory policies towards that group.

Current Research

In the current research, we sought to find an effective intervention to reduce dehumanization of a group consistently dehumanized, namely, Muslims, and explore the underlying mechanism(s) driving the effects of any effective intervention. In Study 1, we utilized an *intervention tournament* design to simultaneously examine the initial efficacy of 10 different interventions, designed in collaboration with Muslim rights practitioners, to address hostility towards Muslims in the United States, vis-à-vis a no-intervention control (for a similar approach, see Bruneau et al., 2018; for a review on intervention tournaments, see Hameiri & Moore-Berg, 2022). We then conducted a follow-up study 1 month later to examine the lasting efficacy of the successful intervention (“Mean Tweets”) identified in Study 1. The Mean Tweets intervention is a 3-minute video of a diverse and relatable group of young Muslims reading aloud, reacting, and laughing at the general public’s Islamophobic comments posted on an online news-coverage video reporting that a Muslim preschool had burnt down in Tampa Bay, Florida.¹ In Study 2, we conducted a preregistered replication of Study 1 to explore the psychological processes driving the effects of the dehumanization-reducing Mean Tweets intervention.

STUDY 1

In Study 1, we examined the efficacy of 10 interventions aimed at reducing hostility towards Muslims. Of these interventions, eight were video interventions, one was an article intervention, and one was an intervention activity highlighting prosocial behavior that has been shown to effectively reduce hostility towards Muslims (Gallardo et al., 2022) (See Table 1 for descriptions and links of all interventions). We partnered with practitioners from the Institute for Social Policy and Understanding (ISPU)—a nonpartisan, Muslim advocacy nonprofit (<https://www.ispu.org/>)—to curate these interventions based on their and our intuitions on what would be an effective intervention to reduce hostility towards Muslims. To test several interventions simultaneously, we used an *intervention tournament* design in which we tested each intervention vis-à-vis a no-intervention control (Hameiri & Moore-Berg, 2022). The study and consent procedures were approved by the Institutional Review Board (IRB) at the University of Pennsylvania.

¹To watch the full video, see <https://player.vimeo.com/video/352558407>.

Table 1. Study 1. Summaries of Each Condition

Condition	Summary of Intervention	Link to Intervention	Length
Control—No intervention	—	—	—
Intervention 1: Mean Tweets (Humor)	A video of Muslims mocking the general public’s Islamophobic comments posted on an online video describing a Muslim preschool that burnt down	https://player.vimeo.com/video/352558407	3:00
Intervention 2: Muslim Women (Statistics)	A statistic-driven video debunking the myth that Muslim women are oppressed	https://player.vimeo.com/video/352962622	1:55
Intervention 3: An America Without Muslims (Statistics)	An informational video describing the ways in which the American society benefits economically and socially from Muslim Americans	https://player.vimeo.com/video/352558389	1:38
Intervention 4: A Muslim Women Explains Why Hijabs Are Not Oppressive (Logical)	A clip from a talk show showcasing a Muslim woman using a logical argument to explain why hijabs are not oppressive	https://player.vimeo.com/video/352962631	0:45
Intervention 5: Muslims Are Non-Violent (Statistics)	A video that presents statistics to counter common misconceptions about Muslims (e.g., Muslims are violent)	https://player.vimeo.com/video/352558365	2:36
Intervention 6: A Muslim Girl Overcomes Her Fear (Empathy)	An empathic video about a girl who overcame her fear of public speaking. At the end of the video, it is revealed that the girl is Muslim	https://player.vimeo.com/video/352558426	1:05
Intervention 7: A Community Helps an Elderly Muslim (Empathy)	A video telling a moving story of a community coming together to walk with an elderly Muslim woman who was harassed on a previous walk	https://player.vimeo.com/video/352558440	1:20
Intervention 8: An American Football Player Turned Muslim (Logical)	A video telling the story of a former football player who despite being once praised for his football skills is now considered a threat for converting to Islam	https://player.vimeo.com/video/352558452	1:20
Intervention 9: Muslims Condemn Terrorism (Logical)	An article about a Muslim schoolgirl who made an extensive list of instances when Muslims condemned terrorism to debunk the common narrative that Muslims do not condemn terrorism	https://www.theguardian.com/world/shortcuts/2017/mar/26/muslims-condemn-terrorism-stats	—
Intervention 10: The Collective Praise Intervention (Hypocrisy-based)	An activity where participants read three different narratives of recent prosocial behaviors performed by a group of Muslims. Then, participants read three different narratives of recent prosocial behaviors performed by a group of Christians. Each vignette was accompanied by a relevant image. After each individual vignette, participants responded on unmarked sliders anchored at 0 (<i>not at all</i>) to 100 (<i>very</i>) to: “How much should the actions of these [Muslim/Christian] organizations reflect upon [Muslims/Christians], in general?”	Gallardo et al. (2022)	—

Methods

Participants

We recruited 2,405 participants via Amazon's Mechanical Turk (Mturk). After removing 30 participants who identified as Muslim and 26 participants who failed an embedded attention check question ("This is an attention check question. Please answer 5 for this question."), a final sample of 2,349 participants remained (53.7% female; 34.8% Republican, 65.2% Democrat; $M_{\text{age}} = 36.8$, $SD = 13.11$; 71.0% White, 11.5% Black, 7.9% Asian, 5.8% Hispanic, 3.8% other; 49.1% Christian, 39.7% atheist/agnostic, 2.0% Judaism, 2.1% Buddhism, 0.9% Hinduism, and 6.2% other), with at least 204 participants per condition. (See [Table S1](#) in the online supporting information for sample size and demographics for each condition). A sensitivity power analysis indicated that when comparing two conditions (i.e., one of our interventions and control), our final sample was sufficiently powered to detect small-medium effects ($d = .36$) with .95 power. Participants were compensated US\$1.50. For the complete survey, syntax, and data, see https://osf.io/95jx4/?view_only=e0dc06e3b6044c7b8306593ce15d53c4.

Measures and Procedure

Participants were randomly assigned to either one of 10 intervention conditions or to the no-intervention control. Participants were asked to complete the intervention (or not in the control), that either included watching a short 45-second to 3-minute video, reading a brief article, or completing a short activity in the Collective Praise Intervention (Gallardo et al., 2022) condition. Following the intervention (or not in the control), all participants completed the key outcome measures of dehumanization and anti-Muslim policy support, as we elaborate on below. Additionally, we included two other exploratory variables (i.e., prejudice, collective blame) that were not the focal point of this article and are thus reported in the online supporting information. Additionally, the NGO we collaborated with included three variables for their internal use (i.e., violence justification, policy support, and Islamophobia index). Finally, we collected demographics.

Dehumanization was assessed with two measures (presentation of each measure relative to the other was randomized): a multi-item trait measure (Bastian & Haslam, 2011) and The Ascent of Hu(Man) scale (Kteily et al., 2015). The multi-item trait measure asked participants how well the following traits (in randomized order) apply to Muslims and Christians (order of presentation of each target group was randomized across participants; R=reverse-coded) using a 7-point scale anchored at 1 (*not at all*) and 7 (*extremely so*): "refined" (R), "cultured" (R), "rational" (R), "logical" (R), "backward," "primitive," "savage," "aggressive," "lacking morals," "barbaric," "cold-hearted," "scientifically/ technologically advanced" (R), "capable of self-control" (R), "mature" (R), and "responsible" (R). We averaged responses to create a score for Muslims ($\alpha = .93$) and a score for Christians ($\alpha = .94$). We then created a trait dehumanization score by computing a difference score between Christians and Muslims. For the second measure of dehumanization, we used a variation of the Ascent of Hu(Man) scale (Kteily et al., 2015) and asked participants to determine where target groups fall on a scale, ranging from 0 (*least evolved*) to 100 (*most evolved*). We assessed "ascent dehumanization" towards Muslims, Christians, and other filler groups (presentation order of all groups were randomized across participants). We assessed ascent dehumanization as a difference score between Muslims and Christians (for a similar approach, see Gallardo et al., 2022). As in previous work by Bruneau and colleagues (e.g., Bruneau et al., 2021; Bruneau & Kteily, 2017; Kteily

et al., 2015), we standardized the ascent dehumanization score and the multi-item trait score, and then we averaged them to provide a single measure of dehumanization ($r = .65, p < .001$).

Anti-Muslim policy support was assessed by asking participants to indicate their support for four policies targeting Muslims (presented in randomized order; R=reverse coded), adapted from Kteily and Bruneau (2017) and Gallardo et al. (2022): “A ban on visas to Muslim wanting to enter the United States”, “A surveillance program targeting mosques in the US,” “Banning the building of mosques in the US,” and “The protection of American Muslims’ civil rights” (R). Responses were made on a 4-point Likert-type scale anchored at 1 (*strongly oppose*) and 4 (*strongly support*). We averaged responses together to create a composite anti-Muslim policy support score ($\alpha = .85$).

Demographics of race, age, gender, religion, and political affiliation were collected.

Results

First, to ensure equal randomization, we conducted a one-way ANOVA with our demographic variables. Results revealed no significant differences in race, religion, age, political affiliation, or gender across conditions ($ps > .059$). Next, we conducted a series of planned contrasts to examine the differences in dehumanization and anti-Muslim policy support between each experimental condition and the no-intervention control. For any experimental condition that significantly reduced dehumanization *and* anti-Muslim policy support (as compared to the no-intervention control), we also tested whether the intervention’s effect on policy support was mediated by dehumanization using Hayes’s (2018) PROCESS (Model 4) (for a similar approach, see Gallardo et al., 2022). For means, standard deviations, and variable intercorrelations, see Table 2. For means and standard deviations by condition, *t*-values, and Cohen’s *d* compared to the control for each condition, see Table 3.

Results indicated that compared to the control ($n = 218$), dehumanization of Muslims ($p = .006, d = .25$) and anti-Muslim policy support ($p = .015, d = .23$) were significantly lower for participants in the Mean Tweets condition ($n = 210$).² There was also a significant indirect effect of the Mean Tweets condition (vs. the control) on anti-Muslim policy support through dehumanization (*indirect effect* = $-.11, SE = .05, 95\%$ confidence interval (CI) = $[-.20, -.03]$; see Figure 1). For all other experimental conditions, as compared to the control, there were no significant difference in dehumanization ($ps > .119, ds < .15$) or in anti-Muslim policy support ($ps > .172, ds < .13$).

One-Month Follow-Up Assessment

We conducted a 1-month follow-up study with participants from the effective Mean Tweets condition and control to examine whether the promising effects of the Mean Tweets video found at Wave 1 persisted. It is worth noting that although the nine other interventions did not have

²As an exploratory analysis, we used Hayes’s (2018) PROCESS macro model 1 to examine whether participants’ religious affiliation moderated the effect of the Mean Tweets condition versus the control on our outcome variables. Given most participants identified as Christian, we recoded our religious affiliation variable such that 1 = *Christians* and 0 = *non-Christians*. Results indicated that there was no significant Condition \times religious affiliation interaction on dehumanization of Muslims ($b = -.11, SE = .17, 95\%$ CI $[-.43, .22]$). However, there was a significant Condition \times religious affiliation interaction on anti-Muslim policy support ($b = -.34, SE = .14, 95\%$ CI $[-.61, -.07]$), such that Christians in the intervention condition displayed significantly lower anti-Muslim policy support versus Christians in the control condition ($b = -.35, SE = .10, 95\%$ CI $[-.54, -.15]$); but not for non-Christian participants ($b = -.01, SE = .10, 95\%$ CI $[-.20, .18]$). Additionally, we found that the Mean Tweets intervention’s effect on dehumanization ($p = .004$) and anti-Muslim policy support ($p = .012$) held when controlling for participants’ religious affiliation.

Table 2. Study 1. Means, Standard Deviations, and Correlations for all Main Variables

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6
1. Dehumanization	.000	.91	—					
2. Anti-Muslim policy support	1.67	.75	.59***	—				
3. Political affiliation	-.93	5.39	.16***	.20***	—			
4. Age	36.80	13.11	.08***	.09***	.20***	—		
5. Gender (M = 1, F = 2)	—	—	-.01	-.01	.52***	.31***	—	
6. Religion (non-Christians = 0, Christians = 1)	.49	.50	.33***	.29***	.17***	.08***	.03	—

Note: *N* = 2349.

*** *p* < .001.

Table 3. Study 1. Means (Standard Deviations) and Planned Contrasts for Key Measures Reported in Main Text

	Condition	Dehumanization	Anti-Muslim Policy Support
<i>M (SD)</i>	Control—No video	.07 (.99)	1.68 (.81)
	Intervention 1: Mean Tweets (Humor)	-.17 (.95)	1.51 (.72)
	Intervention 2: Muslim Women (Statistics)	-.05 (.78)	1.66 (.74)
	Intervention 3: An America Without Muslims (Statistics)	-.06 (.92)	1.62 (.67)
	Intervention 4: A Muslim Woman Explains Why Hijabs Are Not Oppressive (Logical)	.04 (.75)	1.65 (.73)
	Intervention 5: Muslims Are Nonviolent (Statistics)	-.06 (.64)	1.59 (.67)
	Intervention 6: A Muslim Girl Overcomes Her Fear (Empathy)	.16 (1.17)	1.76 (.79)
	Intervention 7: A Community Helps an Elderly Muslim (Empathy)	-.02 (.88)	1.75 (.74)
	Intervention 8: An American Football Player Turned Muslim (Logical)	.09 (1.01)	1.75 (.81)
	Intervention 9: Muslims Condemn Terrorism (Logical)	-.03 (.96)	1.72 (.76)
	Intervention 10: The Collective Praise Intervention (hypocrisy based)	.04 (.80)	1.67 (.71)
	Intervention 1: Mean Tweets (Humor)	2.55 (.25)**	2.35 (.23)*
	Intervention 2: Muslim Women (Statistics)	1.31 (.13)	.38 (.04)
	Intervention 3: An America Without Muslims (Statistics)	1.42 (.14)	.91 (.09)
	Intervention 4: A Muslim Woman Explains Why Hijabs Are Not Oppressive (Logical)	.33 (.03)	.53 (.05)
	Intervention 5: Muslims Are Nonviolent (Statistics)	1.50 (.15)	1.37 (.13)
	Intervention 6: A Muslim Girl Overcomes Her Fear (Empathy)	-.90 (.09)	-.93 (.09)
	Intervention 7: A Community Helps an Elderly Muslim (Empathy)	.95 (.09)	-.88 (.09)
	Intervention 8: An American Football Player Turned Muslim (Logical)	-.21 (.02)	-.88 (.09)
	Intervention 9: Muslims Condemn Terrorism (Logical)	1.05 (.10)	-.41 (.04)
	Intervention 10: The Collective Praise Intervention (hypocrisy based)	.28 (.03)	.20 (.02)

Note: $N = 2,349$.

* $p < .05$; ** $p < .01$.

Planned contrasts comparing target videos and control; t value (Cohen's d)

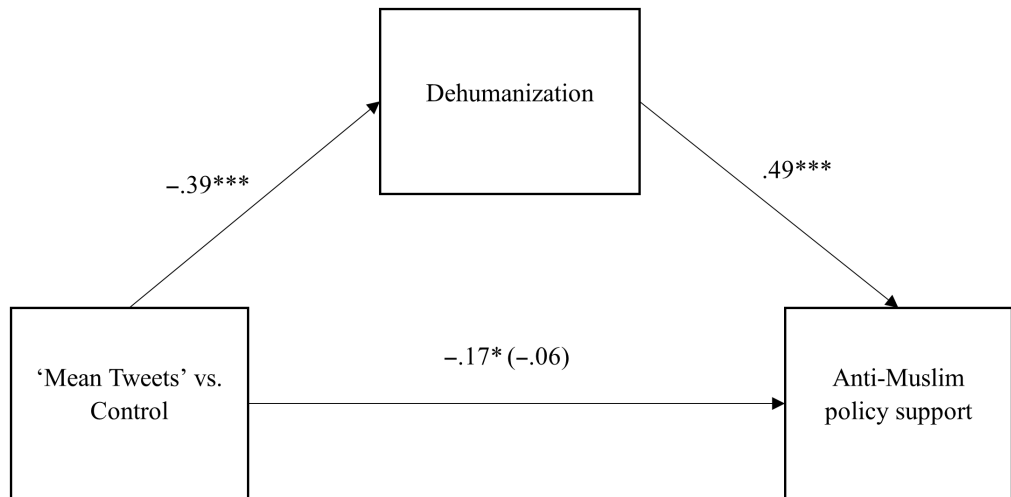


Figure 1. Study 1. Unstandardized path model assessing the effect of the Mean Tweets intervention on anti-Muslim policy support through dehumanization. * $p < .05$; *** $p < .001$.

an immediate effect in reducing dehumanization of Muslims and anti-Muslim policy support compared to the control, there is a possibility that these interventions could have shown an effect at Wave 2 (i.e., sleeper effect; Greenwald, 1975). However, due to budget constraints, we were only able to recruit participants from the control and our most promising intervention, the Mean Tweets intervention. After removing participants who failed an embedded attention check question (“This is an attention check question. Please answer 5 for this question.”), out of the 428 participants from the control and the Mean Tweets condition in Wave 1, 313 participants completed our survey (73% retention; 55.4% female; 36.6% Republican, 62.7% Democrat; $M_{\text{age}} = 37.86$, $SD = 11.48$; 70.1% White, 12.7% Black, 8.3% Asian, 5.7% Hispanic, 2.9% other; 47.8% Christian, 40.5% atheist/agnostic, 2.2% Judaism, 1.9% Buddhism, 1.0% Hinduism, and 6.6% other), with 148 participants in the Mean Tweets condition and 164 in the control (see Table S2 in the online supporting information for demographics for each condition). Attrition ($R^2 = .02$) was not predicted by political affiliation ($p = .854$), ethnicity ($p = .646$), gender ($p = .707$), or condition ($p = .334$) but was predicted by age ($p = .031$): Younger participants were more likely to drop out of the study between Waves 1 and 2. A sensitivity power analysis indicated that when comparing both conditions (i.e., Mean Tweets and control), our final sample was sufficiently powered to detect small-medium effects ($d = .41$) with .95 power. Participants were compensated US\$1.00.

Participants completed the survey without any additional interventions. The survey included the same measures as in Wave 1, including *dehumanization* (multi-item trait dehumanization: Christians: $\alpha = .95$; Muslims: $\alpha = .96$; correlation between ascent dehumanization and multi-item trait dehumanization measures: $r = .78$, $p < .001$) and *anti-Muslim policy support* ($\alpha = .90$), as well as the two other exploratory variables in Wave 1 (i.e., prejudice and collective blame) that are not the focal point of this article. Finally, *demographics* of race, age, gender, religion, and political affiliation were again collected.

We conducted independent samples *t*-tests to examine whether at Wave 2, there was a difference between the Mean Tweets condition and control on dehumanization and anti-Muslim policy

support. Results indicated that when comparing the Mean Tweets and control conditions at Wave 2, dehumanization was descriptively lower, but there was no significant difference ($M_{\text{Control}} = .04$ vs. $M_{\text{Mean Tweets}} = -.04$, $p = .412$, $d = .10$). Additionally, there was no difference in anti-Muslim policy support between the Mean Tweets condition and control at Wave 2 ($M_{\text{Control}} = 1.68$ vs. $M_{\text{Mean Tweets}} = 1.68$, $p = .921$, $d = .01$).

Discussion

Study 1 demonstrated that participants in the Mean Tweets condition versus the control displayed significantly less dehumanization of Muslims and anti-Muslim policy support at Wave 1. Furthermore, there was an indirect effect of the Mean Tweets condition versus the control on anti-Muslim policy support through dehumanization. We reason that showing a diverse and relatable group of young Muslims poking fun at Islamophobic comments led participants in the Mean Tweets condition (vs. the control) to view Muslims as more human (see Haslam, 2006), thereby reducing dehumanization of Muslims and anti-Muslim policy support.

Additionally, even though these effects did not persist 1 month later, a repeated measures ANOVA reported in the online supporting information revealed that the Mean Tweets intervention's effect on dehumanization did not subside across time. It could be that the follow-up study was underpowered to detect such effects. Specifically, a sensitivity analysis indicated that our follow-up study was only sensitive enough to detect an effect size of $d = .41$, while the effect on dehumanization at Wave 2 was $d = .10$. Thus, it is possible that the lack of significant difference on dehumanization at Wave 2 may have been due to a statistical power issue, and if we had more participants at Wave 2, the analyses may have yielded a significant difference. Regarding anti-Muslim policy support, a repeated measures ANOVA revealed that the effect on anti-Muslim policy support simply subsided across time, similar to the effects of the vast majority of short interventions that are longitudinally assessed as part of intervention tournaments (e.g., Bruneau et al., 2022; Lai et al., 2016; Milkman et al., 2021; Moore-Berg et al., 2023; for a review, see Hameiri & Moore-Berg, 2022).

Nevertheless, our intervention tournament revealed that the Mean Tweets video was the only intervention in our intervention tournament to reduce dehumanization towards Muslims and anti-Muslim policy support relative to the control, at least immediately after exposure to the video. However, the current work does not measure an exhaustive list of variables related to hostility towards Muslims (e.g., threat, aggression). Thus, although the nine other interventions failed to reduce dehumanization of Muslims and anti-Muslim policy support, it is worth noting that these interventions may be effective in reducing other variables related to hostility towards Muslims. Additionally, it would be worthwhile to explore why these nine interventions failed to reduce dehumanization of Muslims and anti-Muslim policy support as it would provide useful suggestions into structuring future interventions that seek to reduce these outcomes (see Bar-Tal & Hameiri, 2020). However, the focus of the current work is to identify and explore the mechanism(s) of interventions that effectively reduce dehumanization of Muslims and anti-Muslim policy support. Thus, in Study 2, we aimed to explore the possible mechanisms driving the effects of the Mean Tweets video at Wave 1.

STUDY 2

Study 2 had two goals: (1) to replicate the effects of the successful Mean Tweets video and (2) to understand the possible mechanisms (e.g., humor) driving the effects of the Mean Tweets video. To accomplish our goals, Study 2 had two experimental conditions: the Mean Tweets condition, which was a shortened version of the Mean Tweets video from Study 1, and the "Serious

Mean Tweets” condition, which was a video created from the same source material of the Mean Tweets video but with the humorous reactions and laughing edited out. Thus, the Serious Mean Tweets video only showed the same diverse and relatable group of young Muslims reading aloud and seriously reacting Islamophobic comments posted on the same online news coverage video reporting that a Muslim preschool had burnt down in Tampa Bay, Florida, as in the Mean Tweets video. The purpose of the Serious Mean Tweets condition was to explore whether the video is just as effective at reducing the dehumanization of Muslims and anti-Muslim policy support as the Mean Tweets video when the humorous components are removed. Finally, we included a video control condition—an educational video about Canyonlands National Park in Utah. We opted for a video control condition rather than an empty control to reduce potential selection bias and to reduce the possibility of participants realizing that they are in the control condition, which may then affect how they respond to our self-report questionnaire (Hameiri & Moore-Berg, 2022).

If both experimental videos are equally as effective at reducing dehumanization of Muslims and anti-Muslim policy support, it could be that the effects are driven by exposure to relatable Muslims (Varma, 2020) or parasocial contact (Schiappa et al., 2005), rather than the humorous components of the Mean Tweets video. If so, we argue that removing the humorous reactions and laughing from the video (i.e., Serious Mean Tweets) would not affect the effectiveness of the Mean Tweets video, and in fact, both experimental videos may reduce hostility towards Muslims versus the control—as they both expose participants to Muslim-Americans that are relatable and in a positive light. Accordingly, we examined several exploratory variables to identify other potential mechanisms driving the effects of the successful Mean Tweets video; however, for sake of brevity, we only include the exploratory variables that contribute to the focal point of this article and report all other exploratory variables (e.g., message scrutiny, inclusion of other in the self) in the online supporting information.

Furthermore, previous literature on humor as a persuasion technique suggests that humorous messages can sometimes elicit negative consequences. For example, LaMarre et al. (2014) found that Horatian satire, a light-hearted comedic way of telling the truth, led participants to discount political messages as “just a joke” or “not serious,” thereby short-circuiting the intended impact of the message (see also Nabi et al., 2007). As another example of the negative byproducts of humor, Moyer-Gusé et al. (2011) found that showcasing an unplanned pregnancy in a humorous light on a television program reduced the perceived seriousness of the unplanned pregnancy. Thus, to test whether the Mean Tweets video would elicit similar negative consequences, we introduced three new variables—message discounting (i.e., the notion that the information presented is not relevant to forming judgments about the issues discussed; i.e., Islamophobia; Nabi et al., 2007), perceptions of how serious Muslims perceive Muslim targeted hate speech, and perceived tone of the video. We also included several other exploratory variables (i.e., devaluing the severity of hostility towards Muslims in the United States, scrutiny of the video message) to identify other potential negative consequences of the Mean Tweets video, but these variables were not the focal point of this article and are thus reported in the online supporting information.

As preregistered, our main interest was the comparison between the Mean Tweets condition and the control to examine whether the effects from Study 1 replicate. Specifically, our preregistered hypotheses were: (1) As compared to the control, participants exposed to the Mean Tweets video will express less dehumanization of Muslims and less anti-Muslim policy support. (2) As compared to the control, the Mean Tweets video will have an indirect effect on anti-Muslim policy support through dehumanization. As a secondary and exploratory

research question to understand the mechanisms underlying the effects of the Mean Tweets video, we compared the Serious Mean Tweets condition to the control to examine whether a video with the same content of the Mean Tweets video *without* the humorous reactions and laughing would still reduce hostility towards Muslims. Finally, as exploratory analyses, we compared both experimental conditions (i.e., the Mean Tweets and Serious Mean Tweets conditions) to each other to further explore any differences that might shed some light on the mechanisms underlying the effects of the Mean Tweets and the Serious Mean Tweets videos. The study and consent procedures were approved by the Institutional Review Board (IRB) at the University of Pennsylvania.

Methods

Participants

An a priori power analysis using G*Power 3.1 (Faul et al., 2007) found that when comparing two independent groups, 105 participants per condition are needed to obtain a medium effect ($d=.5$) with .95 power, and a sample of 651 per condition is required to detect a small effect ($d=.2$). To detect small-medium effects, and taking into account budgetary constraints, we opted to recruit 250 participants per condition. Via MTurk, 695 participants completed our survey. After removing seven participants who identified as Muslim and 11 participants who failed an embedded attention check question (“This is a check question. Please choose option 2.”), a final sample of 677 participants remained (36.9% female; 34.4% Republican, 65.5% Democrat; $M_{\text{age}}=37.94$, $SD=11.35$, 64.3% White, 20.1% Black, 8.1% Asian, 4.7% Hispanic, 2.8% other; 68.2% Christian, 26.6% atheist/agnostic, 1.3% Judaism, 1.8% Buddhism, 0.7% Hinduism, and 1.3% other), which according to a sensitivity analysis allowed us to detect small-medium effects ($d=.34$ with .95 power, or $d=.26$ with .80 power) (see Table S3 in the online supporting information for sample size and demographics for each condition). Participants were compensated US\$1.50.

Measures and Procedure

Participants were randomly assigned to one of two experimental video conditions: (1) Mean Tweets ($n=232$) (<https://player.vimeo.com/video/535979120>); (2) Serious Mean Tweets ($n=221$) (<https://player.vimeo.com/video/535979239>); or to the control video ($n=224$) (<https://player.vimeo.com/video/542206323>). To ensure that video length did not influence the video’s effects, we edited each video to be similar in length (between 1:33 and 1:38 minutes). Following the video, all participants completed a humor manipulation check, the same dehumanization measure as in Study 1, a revised measure of anti-Muslim policy support, tone, perceptions of how serious Muslims consider Muslim targeted hate speech, and a message discounting measure. Study 2 was preregistered before data collection; see https://osf.io/95jx4/?view_only=2b817d42e529471a9c58a7ff23591234. For the complete survey, syntax, and data, see https://osf.io/95jx4/?view_only=e0dc06e3b6044c7b8306593ce15d53c4.

The manipulation check was assessed by asking participants to indicate their agreement that the video was “Amusing” and “Funny” (and other filler items in random order). Responses were made on a 7-point Likert-type scale anchored at 1 (*not at all*) and 7 (*very much*). We averaged

and combined responses to “Amusing” and “Funny” to create the manipulation check score ($r = .67, p < .001$).

Dehumanization was assessed as in Study 1 (multi-item trait dehumanization: Christians: $\alpha = .88$; Muslims: $\alpha = .90$; correlation between ascent dehumanization and multi-item trait dehumanization measures: $r = .49, p < .001$).

Anti-Muslim policy support was assessed by asking participants to indicate their support for seven policies targeting Muslims taken from Kteily and Bruneau (2017) (for a similar approach, see Gallardo et al., 2022). Participants rated their support on the following seven policies (in randomized order; R=reverse coded): “We need to stop accepting Muslim refugees into this country, period”; “No more political correctness: Muslim refugees need to go”; “We should stop giving out visas to Muslims to enter the U.S.”; “We should tap the phones of any Muslims in the U.S. who make calls to Middle Eastern countries on a frequent basis”; “We should ban the opening of any new Mosques in this country”; “We should ban the wearing of the Islamic veil (or ‘headdress’)”; and “We should allow Muslims to practice their religion with no restrictions in the U.S.” (R). Responses for all items were made on a 7-point Likert-type scale anchored at 1 (*strongly disagree*) and 7 (*strongly agree*) and averaged together to create an anti-Muslim policy support score ($\alpha = .93$).

Tone was assessed by asking participants to respond to the following statement: “Using the slider below, please indicate how light-hearted or critical/harsh the tone of the video was.” Responses were made using an unmarked slider bar anchored at 0 (*Light-Hearted*) and 100 (*Critical/Harsh*).

Perception of How Serious Muslims Perceive Muslim Targeted Hate Speech was assessed by asking participants to indicate how serious Muslims think the following five forms of hate speech committed towards Muslims are: “insults,” “threats,” “hate mail or emails,” “hateful language posted online or on social media,” and “bullying.” Responses were made using unmarked slider bars anchored at 0 (*not at all serious*) and 100 (*extremely serious*) and averaged together to create the perceived Muslim perception of Muslim targeted hate-speech score ($\alpha = .90$).

Message discounting was assessed by asking participants to indicate their agreement with four items adapted from Nabi et al. (2007) (in randomized order; R=reverse coded): “The author of the message was just joking”; “The message was intended more to entertain than to persuade”; “The author was serious about advancing his views in the message” (R); and “It would be easy to dismiss this message as simply as joke.” Responses were made on a 7-point Likert-type scale anchored at 1 (*strongly disagree*) and 7 (*strongly agree*) and averaged together to create a message discounting score ($\alpha = .73$).

Demographics were collected as in Study 1.

Results

As in Study 1, we first conducted a one-way ANOVA with our demographic variables to ensure equal randomization. Results indicated no significant differences in race, religion, age, political affiliation, or gender across conditions ($ps > .145$). Next, to ensure that the manipulation worked as intended, we examined whether the Mean Tweets video was indeed more humorous than the control and the Serious Mean Tweets video. A one-way ANOVA indicated that there was a main effect of condition on our manipulation check, humor, $F(2, 674) = 28.00, p < .001$, partial $\eta^2 = .077$. Post hoc comparisons revealed that participants in the Mean Tweets condition reported the video was significantly *more* humorous than the control and the Serious Mean

Tweets video ($p < .001$, $d = .44$; $p < .001$, $d = .70$, respectively), while participants in the Serious Mean Tweets condition reported the video was significantly *less* humorous as compared to the control ($p = .002$, $d = .28$). For means, standard deviations, and variable intercorrelations, see Table 4.

Preregistered Main Analyses

As preregistered, we then conducted a series of planned comparisons to examine the differences between the experimental conditions versus the control on our main dependent variables: dehumanization and anti-Muslim policy support. For means and standard deviations by condition and t -values and Cohen's d compared to the control for each condition, see Table 5. Consistent with Study 1 and as preregistered, we also tested whether the intervention affected anti-Muslim policy support through dehumanization using Hayes's (2018) PROCESS (Model 4).

Compared to the control, participants in the Mean Tweets condition dehumanized Muslims significantly less ($p < .001$, $d = .32$),³ but there was no difference between these two conditions on anti-Muslim policy support ($p = .508$, $d = .06$). However, there was a significant indirect effect of the Mean Tweets condition (vs. the control) on anti-Muslim policy support through dehumanization (*indirect effect* = $-.22$, $SE = .06$, 95% CI = $[-.35, -.10]$; see the online supporting information).

Similarly, compared to the control, participants in the Serious Mean Tweets condition dehumanized Muslims significantly less ($p = .001$, $d = .31$), but there was no difference between these two conditions on anti-Muslim policy support ($p = .800$, $d = .02$). There was also a significant indirect effect of the Serious Mean Tweets condition (vs. the control) on anti-Muslim policy support through dehumanization (*indirect effect* = $-.21$, $SE = .06$, 95% CI = $[-.33, -.10]$; see the online supporting information).

Preregistered Exploratory Analyses

To further explore our results, we conducted additional post hoc analyses comparing the Mean Tweets condition to the Serious Mean Tweets condition on all variables (i.e., dehumanization, anti-Muslim policy support, tone, perception of how serious Muslims perceive Muslim targeted hate speech, and message discounting). A one-way ANOVA revealed a main effect of condition on dehumanization, $F(2, 674) = 7.45$, $p < .001$, partial $\eta^2 = .022$. Post hoc comparisons revealed that between participants in the Serious Mean Tweets and Mean Tweets condition, there was no significant difference in dehumanization of Muslims ($p = .950$, $d = .03$). Additionally, a one-way ANOVA revealed no main effect of condition on anti-Muslim policy support ($F(2, 674) = .22$, $p = .800$, partial $\eta^2 = .001$).

Next, looking at tone, a one-way ANOVA revealed a main effect of condition, $F(2, 674) = 53.71$, $p < .001$, partial $\eta^2 = .138$. Post hoc comparisons revealed that as compared to the

³As in Study 1, we again used Hayes's (2018) PROCESS macro model 1 to examine whether participants' religious affiliation moderated the effect of the Mean Tweets condition versus the control on our outcome variables. Consistent with Study 1, we recoded our religious affiliation variable such that 1 = *Christians* and 0 = *non-Christians*. Results indicated that there were no significant Condition \times religious affiliation interactions on dehumanization of Muslims ($b = .09$, $SE = .16$, 95% CI $[-.23, .40]$) or on anti-Muslim policy support ($b = -.32$, $SE = .31$, 95% CI $[-.93, .29]$). As in Study 1, the Mean Tweets intervention's effect on dehumanization holds when controlling for participants' religious affiliation ($p < .001$).

Table 4. Study 2. Means, Standard Deviations, and Correlations for All Main Variables

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9
1. Humor	3.75	1.92	—								
2. Dehumanization	-.001	.86	.10*	—							
3. Anti-Muslim policy support	3.29	1.75	.50***	.39***	—						
4. Tone	54.52	30.94	.02	-.01	.31***	—					
5. Perceptions of how serious Muslims perceive Muslim hate speech	73.05	19.47	-.21***	-.17***	-.23***	.09*	—				
6. Discounting message	3.35	1.44	.62***	.19***	.66***	.17***	-.22***	—			
7. Political affiliation	-.73	2.44	.05	.26***	.19***	.08	-.06	.04	—		
8. Age	37.94	11.35	-.17***	-.02	-.10*	-.04	.16***	-.15***	.13***	—	
9. Gender (<i>M</i> = 1, <i>F</i> = 2)	—	—	-.19***	-.06	-.04	-.01	.15***	-.09*	-.09*	.11**	—
10. Religion (non-Christians = 0, Christians = 1)	.68	.47	.34***	.33***	.48***	.23***	.14***	.40***	.15***	.02	.02

Note: *N* = 677.

p* < .05; *p* < .01; ****p* < .001.

Table 5. Study 2. Means (Standard Deviations) and Planned Contrasts for Key Measures Reported in Main Text

M (SD)	Condition	Humor	Dehumanization	Anti-Muslim Policy Support	Tone	Perceptions of How Serious Muslims Targeted	
						Muslim Targeted Hate Speech	Discounting Message
	Control Video	3.67 (1.83)	.17 (.78)	3.34 (1.72)	44.95 (29.55)	71.90 (19.85)	3.40 (1.29)
	Intervention 1: Mean Tweets	4.44 (1.70)	-.10 (.91)	3.23 (1.75)	48.14 (31.43)	70.91 (20.29)	3.57 (1.44)
	Intervention 2: Serious Mean Tweets	3.14 (2.00)	-.09 (.87)	3.29 (1.79)	70.75 (24.95)	76.35 (17.84)	3.09 (1.53)
	<i>Planned contrasts comparing target videos and control: t value (Cohen's d)</i>						
	Intervention 1: Mean Tweets	4.46 (.44)***	-3.36 (.32)**	-.66 (.06)	1.18 (.10)	-.54 (.05)	1.26 (.12)
	Intervention 2: Serious Mean Tweets	-3.05 (.28)**	-3.30 (.31)**	-.25 (.02)	9.55 (.94)***	2.45 (.24)*	-2.33 (.22)*

Note: N = 677.

*p < .05; **p < .01; ***p < .001.

control, there was no difference in tone in the Mean Tweets condition ($p = .239$, $d = .10$); however, participants in the Serious Mean Tweets condition reported the tone of the video significantly *more* harsh than participants in the control ($p < .001$, $d = .94$). Additionally, compared to the Serious Mean Tweets condition, participants in the Mean Tweets condition reported the tone significantly *less* harsh ($p < .001$, $d = .74$).

Next, looking at perception of how serious Muslims perceive Muslim targeted hate speech, a one-way ANOVA revealed a main effect of condition, $F(2, 674) = 4.99$, $p = .007$, partial $\eta^2 = .015$. Post hoc comparisons revealed that as compared to the control, there was no difference in the perception of how serious Muslims perceive Muslim targeted hate speech in the Mean Tweets condition ($p = .590$, $d = .05$); however, participants in the Serious Mean Tweets condition reported that Muslims perceive Muslim targeted hate speech *more* serious ($p = .015$, $d = .24$). Additionally, compared to the Serious Mean Tweets condition, participants in the Mean Tweets reported that Muslims perceive Muslim targeted hate speech *less* serious ($p = .003$, $d = .28$).

Finally, looking at message discounting, a one-way ANOVA revealed a main effect of condition, $F(2, 674) = 6.50$, $p = .002$, partial $\eta^2 = .019$. Post hoc comparisons revealed that as compared to the control, there was no difference in message discounting in the Mean Tweets condition ($p = .207$, $d = .22$); however, participants in the Serious Mean Tweets condition discounted the message of the Serious Mean Tweets video significantly less than the message of the control video ($p = .020$, $d = .22$). More importantly, compared to the Serious Mean Tweets condition, participants in the Mean Tweets condition discounted the message of the Mean Tweets video significantly *more* ($p < .001$, $d = .32$).

In Study 2, to better understand the lack of effects of our interventions on anti-Muslim policy support (as compared to the control), we conducted mediation analyses using Hayes's (2018) PROCESS (Model 4) with Helmert coding. Since both experimental conditions affected dehumanization similarly, we examined the indirect effect of dehumanization (as well as our additional mediating variables, that is, tone, perceptions of how serious Muslims perceive Muslim targeted hate speech, and message discounting) on the relationship between both experimental conditions collapsed (vs. the control) and anti-Muslim policy support. Results showed a significant indirect effect of the collapsed experimental conditions (vs. the control) on anti-Muslim policy support through dehumanization (*indirect effect* = $-.15$, $SE = .04$, 95% CI = $[-.23, -.08]$; see Figure 2). Additionally, our model revealed that there was a significant indirect effect through tone (*indirect effect* = $.10$, $SE = .02$, 95% CI = $[.06, .15]$) but not through perceptions of how serious Muslims perceive Muslim targeted hate speech (*indirect effect* = $-.01$, $SE = .01$, 95% CI = $[-.02, .01]$) or message discounting (*indirect effect* = $-.03$, $SE = .04$, 95% CI = $[-.12, .06]$). The significant indirect of tone when comparing our collapsed experimental conditions versus the control on anti-Muslim policy is likely driven by the Serious Mean Tweets condition, as it was perceived as more harsh in tone compared to both the Mean Tweets and the control conditions; while, there was no difference in perceived tone between the Mean Tweets and the control.

Additionally, since both experimental conditions had diverging effects on tone, perceptions of how serious Muslims perceive Muslim targeted hate speech and message discounting, but not on dehumanization, in the same model, we examined the indirect effect of our four mediating variables on the relationship between the Mean Tweets video (vs. the Serious Mean Tweets video) and anti-Muslim policy support. Results indicated a significant indirect effect of the Mean Tweets condition (vs. the Serious Mean Tweets condition) through tone (*indirect effect* = $.28$, $SE = .05$, 95% CI = $[.18, .39]$), perceptions of how serious Muslims perceive Muslim

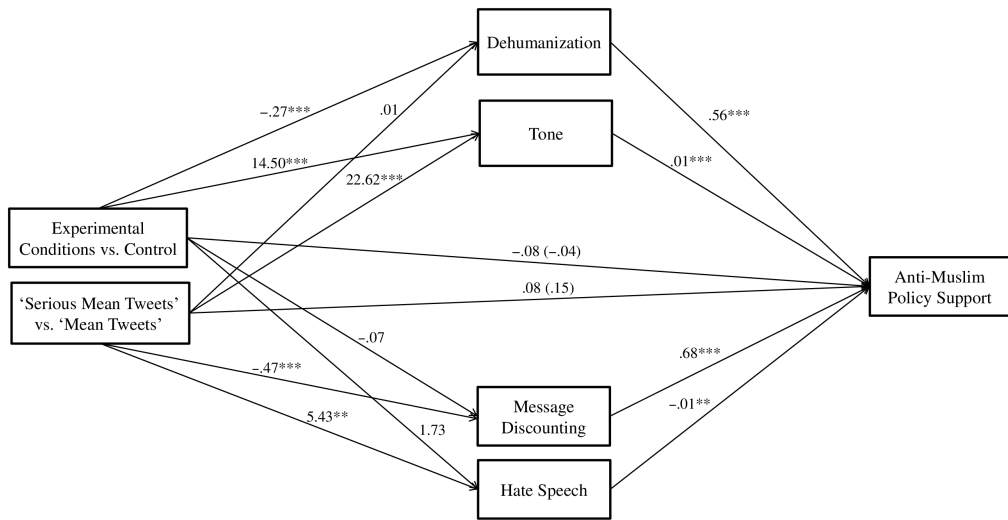


Figure 2. Study 2. Unstandardized path model assessing the effect of (a) the combined experimental conditions versus the control on anti-Muslim policy support through dehumanization and (b) the Mean Tweets condition versus the Serious Mean Tweets condition on anti-Muslim policy support through tone, perceptions of how serious Muslims perceive Muslim targeted hate speech, and message discounting. *** $p < .001$.

targeted hate speech (*indirect effect* = $-.04$, $SE = .02$, 95% $CI = [-.09, -.01]$), and message discounting (*indirect effect* = $-.32$, $SE = .10$, 95% $CI = [-.51, -.13]$), but not through dehumanization (*indirect effect* = $.002$, $SE = .03$, 95% $CI = [-.05, .05]$; see Figure 2).

Mini Meta-Analyses

We conducted two mini meta-analyses (Goh et al., 2016) using fixed effects in which the mean effect size was weighted by sample size to measure the overall effect of the Mean Tweets condition versus the control on dehumanization and anti-Muslim policy support across Studies 1 and 2. The mini meta-analyses showed that the effects across studies were significant for both dehumanization (mean $d = .29$, $Z = 4.22$, $p < .001$, two-tailed) and anti-Muslim policy support (mean $d = .14$, $Z = 2.11$, $p = .035$, two-tailed).

Discussion

Consistent with Study 1, and as preregistered, Study 2 revealed that as compared to a control video, the Mean Tweets video significantly reduced dehumanization towards Muslims and had a significant indirect effect on anti-Muslim policy support through dehumanization. Also, although the Mean Tweets video did not have a significant direct effect on anti-Muslim policy support, a mini meta-analysis confirmed that the Mean Tweets video significantly reduced anti-Muslim policy support across both studies, although the overall effect size was small.

Similarly, as compared to the control, the Serious Mean Tweets video also significantly reduced dehumanization towards Muslims and had a significant indirect effect on anti-Muslim policy support through dehumanization. The humorous Mean Tweets video and the nonhumorous

Serious Mean Tweets video, as indicated by our manipulation check of humor, showed similar effects in reducing hostility towards Muslims. This provides some evidence that it is not only humor that is driving the effects of the Mean Tweets video. Indeed, it is possible that the effects of the Mean Tweets and Serious Mean Tweets videos were at least partly driven by psychological processes associated with the relatability of the Muslims in the video. That is, seeing relatable Muslims act similarly to the participants humanized Muslims in general (e.g., Varma, 2020). It is possible that both interventions worked through parasocial contact—positive media portrayal of Muslims (see e.g., Bruneau et al., 2022). We will discuss these possibilities further in the general discussion.

Furthermore, to further explore the Mean Tweets video's effect on anti-Muslim policy support and the mechanisms driving the effects of the Mean Tweets video, we tested for any differences between the Mean Tweets video and the Serious Mean Tweets video. Results indicated that as compared to participants who watched the Serious Mean Tweets video, participants who watched the Mean Tweets video felt that the tone of the video was significantly less harsh, reported that Muslims perceive Muslim hate speech significantly less serious, which is consistent with previous work, and discounted the video's message significantly more, which is also consistent with previous work (e.g., LaMarre et al., 2014; Moyer-Gusé et al., 2011; Nabi et al., 2007). This provides some evidence that an intervention with a humorous undertone could have some negative consequences.

In exploratory analyses, mediation analyses revealed that while the Mean Tweets video humanized Muslims, which in turn reduced anti-Muslim policy support, the Mean Tweets video also led participants to perceive the video tone as less harsh, feel that Muslims perceive Muslim targeted hate speech less seriously, and to greater message discounting (compared to the Serious Mean Tweets video), which all, in turn, were associated with anti-Muslim policy support. These diverging effects might explain the small and inconsistent effect of the Mean Tweets video on anti-Muslim policy support. Additionally, when compared to the control, the Serious Mean Tweets video led participants to perceive the tone as harsher, indicate that Muslims perceive Muslim targeted hate speech more seriously, and discount the message in the video far less. Finally, the Serious Mean Tweets video was also discounted far less compared to the Mean Tweets video, and similar to the Mean Tweets video, it did not have an effect on anti-Muslim policy support compared to the control. We will discuss the (lack of) effects on anti-Muslim policy support further next.

General Discussion

Designing and assessing interventions to mitigate the dehumanization of Muslims is an important endeavor given the widespread modern dehumanization of Muslims and the associated detrimental consequences (e.g., Bruneau & Kteily, 2017; Kteily & Bruneau, 2017; Montiel et al., 2019). In the current research, utilizing an intervention tournament that simultaneously tested 10 different interventions aimed at reducing hostility towards Muslims, we found that a brief video (i.e., Mean Tweets) that showcased a group of diverse and relatable Muslims reading aloud, reacting, and ridiculing the general public's Islamophobic comments posted on an online news-coverage video reporting that a Muslim preschool had burnt down reduced the dehumanization of Muslims, and, in turn, anti-Muslim policy support. Although a follow-up assessment revealed that these effects did not persist 1 month later, in a preregistered replication study (i.e., Study 2), we found that the effects of the Mean Tweets intervention were partially replicated. Specifically, the Mean Tweets intervention again reduced the dehumanization of Muslims, and although it did not have a direct effect on anti-Muslim policy support, the Mean Tweets video indirectly reduced anti-Muslim policy support through dehumanization. Furthermore, a mini meta-analysis revealed that across both

studies, the Mean Tweets video reduced both dehumanization of Muslims and anti-Muslim policy support. Furthermore, while exploring the lack of direct effect on anti-Muslim policy support in Study 2, we found that our intervention had unintended consequence: It led (at least some of) the participants to discount the intervention's intended message as "just a joke," to report that Muslims perceive Muslim targeted hate speech less serious, and to perceive the tone as less harsh. We suggest that these unintended effects might have contributed to the Mean Tweets intervention's rather weak and inconsistent effects on anti-Muslim policy support.

Additionally, in Study 2, we also found that a version of the Mean Tweets video without the humorous reactions and laughing (i.e., Serious Mean Tweets video), as compared to a control, also reduced dehumanization of Muslims. This suggests that humor is not the sole ingredient driving the effects of the Mean Tweets video, and it is likely that there are other variables driving our effects. For example, it is possible that parasocial contact, or the positive media portrayal of Muslims, could be partly driving the reduction of dehumanization of Muslims in both the Mean Tweets and Serious Mean Tweets videos. Previous work has found that positive media portrayals of outgroups improve intergroup relations (Dovidio et al., 2011; Schiappa et al., 2005), and both videos had positive media portrayals of Muslims. Additionally, in Study 1, it is possible that participants in the other interventions subtyped the Muslims portrayed in the videos (Liskov & Wing, 1994). That is, participants felt that the Muslims in the interventions were exceptions and not representative of the general Muslim population. This could further explain why both the Mean Tweets and Serious Mean Tweets videos were effective in reducing dehumanization of Muslims, as these videos exposed participants to a diverse group of Muslims making subtyping less likely to happen. Relatedly, another possible mechanism driving the effects of both the Mean Tweets and the Serious Mean Tweets interventions could be the relatability of the Muslims portrayed in the videos. Previous work has found that depicting a dehumanized group as relatable can increase the attributed humanity towards the dehumanized group (e.g., Honary et al., 2018; Varma, 2020).

Finally, it could be that humor is still partly responsible for our effects. Humor is a persuasive trait—it increases motivation to process a message (Zhang & Zinkhan, 2006); it attracts attention over nonhumorous messages (Blanc & Brigaud, 2014); it reduces negative cognitions related to the message; and it increases positive affect (Cline & Kellaris, 2007). Furthermore, there is reason to believe that using humor as an intervention technique by displaying a dehumanized group as humorous may be particularly effective in reducing the dehumanization of that group and warrants another reason to examine humor's potential in reducing dehumanization. Indeed, the literature on dehumanization states that denying humans uniquely human traits is a form of dehumanization (i.e., animalistic dehumanization; Haslam, 2006), while the literature on humor argues that humor is a uniquely human trait (Martin & Ford, 2018). Thus, showing a group using humor might increase people's perceptions of their humanity. Ultimately, we argue that our work contributes to the literature on dehumanization by exploring the mechanisms of a dehumanization-reducing intervention and finding that one video reliably reduced the dehumanization of Muslims and indirectly reduced hostile policy support towards Muslims through dehumanization.

We also reason that our intervention's effects contribute to the literature on satire. In Study 1, the Mean Tweets intervention reduced anti-Muslim policy support, albeit with small effects, but in Study 2, it evidenced no effect on anti-Muslim policy support. To better understand this inconsistent effect on anti-Muslim policy support, we conducted exploratory analyses and found that participants who watched the Mean Tweets video discounted the video message as "just a joke," felt Muslims perceive Muslims targeted hate speech less

serious, and reported the tone as less harsh (as compared to participants who watched a nonhumorous video created from the same source material as the Mean Tweets video, i.e., Serious Mean Tweets). Moreover, in the Mean Tweets video, we found that the more participants discounted the message, reported that Muslims perceive Muslim targeted hate speech less serious, and reported that the tone was less harsh, the more they supported anti-Muslim policies. Thus, we argue that it is possible that the Mean Tweets intervention's unintended effects could have contributed to the short-term and inconsistent effects of the Mean Tweets intervention on the outcome measures. This finding corresponds to Nabi et al.'s (2007) research on political humor. Nabi et al. found that participants exposed to a political message embedded in satirical humor discounted the arguments of the message and consequently did not consider the arguments of the message relevant when forming their political opinion (see also Innocenti & Miller, 2016; LaMarre et al., 2014). Therefore, our results add to the literature on satire by finding a similar pattern with humor in the context of intergroup relations and suggest that humor, and in particular satirical humor, as a persuasion tool, regardless of context, may have a negative effect.

Finally, we argue that empirically testing a parasocial contact intervention—a form of intergroup contact that displays positive portrayal of the outgroup in the media—also contributes to the literature on intergroup contact. The existing literature states that the content of intergroup contact interventions is key to the intervention's effectiveness, as contact in and of itself is not always enough to improve intergroup attitudes and can even have adverse effects when the contact is negative (e.g., IJzerman et al., 2020; Saguy, 2018). In fact, Barlow et al. (2012) found evidence across two studies that negative intergroup contact is a stronger and more consistent predictor of negative intergroup outcomes (i.e., increased discrimination) than positive contact is of positive intergroup outcomes. Thus, we argue that it is crucial to empirically test the content of intergroup contact interventions, as well-intended intergroup contact interventions may sometimes have negative intergroup outcomes (e.g., IJzerman et al., 2020; Saguy, 2018). Our research, therefore, contributes to the literature on intergroup contact by revealing that our video depicting a relatable and diverse group of Muslims in a parasocial contact intervention can effectively humanize a disadvantaged group.

Additionally, given that the Mean Tweets video had unintended consequences (e.g., message discounting), we suggest that if researchers or practitioners are interested in utilizing humorous interventions, that they carefully implement them. One approach to counteracting these potential negative consequences could be by ending the humorous messages with serious content, known as the “restoration of gravity” approach (see Nabi et al., 2007). For example, in the Mean Tweets intervention, we could have concluded the humorous video with a serious statement (e.g., stating that the tweets highlight the current hostility towards Muslims and calling for others to help combat hostility towards Muslims). According to Nabi et al. (2007), this “restoration of gravity” approach may allow the Mean Tweets intervention to retain the persuasive benefits of the humorous message (e.g., positive affect) while reducing the associated negative effects. We encourage future research to test this “restoration of gravity” approach in humorous interventions designed to improve intergroup relations.

Despite these key findings and contributions, our work is not without limitations. One limitation is that we cannot decisively report the mechanism driving the effects of the dehumanization-reducing Mean Tweets intervention. Although there are elements of positive parasocial contact, relatability with Muslims, and humor, we cannot conclusively state that any one of these elements are solely driving our effects. Nevertheless, we were able to

identify and better understand an intervention that reliably reduced the dehumanization of Muslims.

In summary, the current work provides evidence that an intervention displaying Muslims reacting, mocking, and laughing at Islamophobic comments reliably reduces the dehumanization of Muslims and indirectly reduces anti-Muslim policy support through dehumanization. Additionally, we discovered unintended consequences of our intervention: Our intervention led participants to discount the intervention's intended message and report that Muslims perceive Muslim targeted hate speech less seriously. We speculate these unintended consequences could partially explain the intervention's inconsistent direct effects on anti-Muslim policy support. Ultimately, we suggest that the current research provides valuable contributions to the literature on dehumanization, intergroup contact, and satire.

ACKNOWLEDGMENTS

We are grateful to Emile Bruneau for inspiration; the general idea in this article reflects conversations and collaborations with Emile in his effort to put science to work for peace, and we were deeply saddened by his loss to brain cancer on September 30, 2020. Correspondence concerning this article should be addressed to Roman A. Gallardo, Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA. E-mail: rgallard@chicagobooth.edu and Boaz Hameiri, The Evens Program in Conflict Resolution and Mediation, Tel Aviv University, Tel Aviv, Israel. E-mail: bhameiri@tauex.tau.ac.il

REFERENCES

- Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.
- Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R., Harwood, J., Rubin, M., & Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, 38(12), 1629–1643. <https://doi.org/10.1177/0146167212457953>
- Bar-Tal, D., & Hameiri, B. (2020). Interventions to change well-anchored attitudes in the context of intergroup conflict. *Social and Personality Psychology Compass*, 14(7), e12534. <https://doi.org/10.1111/spc3.12534>
- 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>
- Beard, F. K. (2005). One hundred years of humor in American advertising. *Journal of Macromarketing*, 25(1), 54–65. <https://doi.org/10.1177/0276146705274965>
- Blanc, N., & Brigaud, E. (2014). Humor in print health advertisements: Enhanced attention, privileged recognition, and persuasiveness of preventive messages. *Health Communication*, 29(7), 669–677. <https://doi.org/10.1080/10410236.2013.769832>
- Bruneau, E., Casas, A., Hameiri, B., & Kteily, N. (2022). Exposure to a media intervention helps promote support for peace in Colombia. *Nature Human Behaviour*, 6, 847–857. <https://doi.org/10.1038/s41562-022-01330-w>
- Bruneau, E., Hameiri, B., Moore-Berg, S. L., & Kteily, N. (2021). Intergroup contact reduces dehumanization and meta-dehumanization: Cross-sectional, longitudinal, and quasi-experimental evidence from 16 samples in five countries. *Personality and Social Psychology Bulletin*, 47(6), 906–920. <https://doi.org/10.1037/pspp0000048>
- Bruneau, E., & Kteily, N. (2017). The enemy as animal: Symmetric dehumanization during asymmetric warfare. *PLoS One*, 12(7), e0181422. <https://doi.org/10.1371/journal.pone.0181422>
- Bruneau, E., Kteily, N., & Falk, E. (2018). Interventions highlighting hypocrisy reduce collective blame of Muslims for individual acts of violence and assuage anti-Muslim hostility. *Personality and Social Psychology Bulletin*, 44(3), 430–448. <https://doi.org/10.1177/0146167217744197>

- Bruneau, E. G., Kteily, N. S., & Urbiola, A. (2020). A collective blame hypocrisy intervention enduringly reduces hostility towards Muslims. *Nature Human Behaviour*, 4(1), 45–54. <https://doi.org/10.1038/s41562-019-0747-7>
- Capozza, D., Falvo, R., Di Bernardo, G. A., Vezzali, L., & Visintin, E. P. (2014). Intergroup contact as a strategy to improve humanness attributions: A review of studies. *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, 21(3), 349–362.
- Capozza, D., Falvo, R., Trifiletti, E., & Pagani, A. (2014). Cross-group friendships, extended contact, and humanity attributions to homosexuals. *Procedia-Social and Behavioral Sciences*, 114, 276–282. <https://doi.org/10.1016/j.sbspro.2013.12.698>
- Cline, T. W., & Kellaris, J. J. (2007). The influence of humor strength and humor—message relatedness on ad memorability: A dual process model. *Journal of Advertising*, 36(1), 55–67. <https://doi.org/10.2753/JOA0091-3367360104>
- Cooper, C. D. (2005). Just joking around? Employee humor expression as an ingratiation behavior. *Academy of Management Review*, 30(4), 765–776. <https://doi.org/10.5465/amr.2005.18378877>
- Dovidio, J. F., Eller, A., & Hewstone, M. (2011). Improving intergroup relations through direct, extended and other forms of indirect contact. *Group Processes & Intergroup Relations*, 14(2), 147–160. <https://doi.org/10.1177/1368430210390555>
- Fata, I. A., Komariah, E., & Irfandy, I. (2018). Laugh and learn: Evaluating from students' perspective of humor used in English Class. *Al-Ta'lim Journal*, 25(2), 117–127. <https://doi.org/10.15548/jt.v25i2.425>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BRM.41.4.1149>
- Gallardo, R. A., Hameiri, B., Moore-Berg, S. L., & Bruneau, E. (2022). The collective praise intervention: A brief intervention highlighting prosocial behavior reduces hostility towards Muslims. *Group Processes & Intergroup Relations*, 25(7), 1696–1717. <https://doi.org/10.1177/13684302211021665>
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, 10(10), 535–549. <https://doi.org/10.1111/spc3.12267>
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82(1), 1–20. <https://doi.org/10.1037/h0076157>
- Hameiri, B., & Moore-Berg, S. L. (2022). Intervention tournaments: An overview of concept, design, and implementation. *Perspectives on Psychological Science*, 17, 1525–1540. <https://doi.org/10.1177/17456916211058090>
- Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4
- Haslam, N. (2015). Dehumanization and intergroup relations. In M. Mikulincer, P. R. Shaver, J. F. Dovidio, & J. A. Simpson (Eds.), *APA handbook of personality and social psychology, Vol. 2. Group processes* (pp. 295–314). American Psychological Association. <https://doi.org/10.1037/14342-011>
- Haslam, N., & Stratemeyer, M. (2016). Recent research on dehumanization. *Current Opinion in Psychology*, 11, 25–29. <https://doi.org/10.1016/j.copsyc.2016.03.009>
- Hay, A. C., Sheppes, G., Gross, J. J., & Gruber, J. (2015). Choosing how to feel: Emotion regulation choice in bipolar disorder. *Emotion*, 15(2), 139–145. <https://doi.org/10.1037/emo0000024>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.
- Honary, M., McNaney, R., & Lobban, F. (2018). Designing video stories around the lived experience of severe mental illness. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction* (pp. 25–38). <https://doi.org/10.1145/3240167.3240188>
- Ijzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, 4(11), 1092–1094. <https://doi.org/10.1038/s41562-020-00990-w>
- Innocenti, B., & Miller, E. (2016). The persuasive force of political humor. *Journal of Communication*, 66(3), 366–385. <https://doi.org/10.1111/jcom.12231>
- Jardina, A., & Piston, S. (2021). The effects of dehumanizing attitudes about Black people on whites' voting decisions. *British Journal of Political Science*, 52(3), 1076–1098. <https://doi.org/10.1017/S0007123421000089>
- Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology*, 109(5), 901–931. <https://doi.org/10.1037/pspp0000048>

- Kteily, N., Hodson, G., & Bruneau, E. (2016). They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. *Journal of Personality and Social Psychology*, *110*(3), 343–370. <https://doi.org/10.1037/pspa0000044>
- Kteily, N. S., & Bruneau, E. (2017). Darker demons of our nature: The need to (re) focus attention on blatant forms of dehumanization. *Current Directions in Psychological Science*, *26*(6), 487–494. <https://doi.org/10.1177/0963721417708230>
- Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M., Burns, M., ... Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*, 1001–1016. <https://doi.org/10.1037/xge0000179>
- LaMarre, H. L., Landreville, K. D., Young, D., & Gilkerson, N. (2014). Humor works in funny ways: Examining satirical tone as a key determinant in political humor message processing. *Mass Communication and Society*, *17*(3), 400–423. <https://doi.org/10.1080/15205436.2014.891137>
- Liskov, B. H., & Wing, J. M. (1994). A behavioral notion of subtyping. *ACM Transactions on Programming Languages and Systems*, *16*(6), 1811–1841. <https://doi.org/10.1145/197320.197383>
- Martin, R. A., & Ford, T. (2018). *The psychology of humor: An integrative approach*. Academic Press.
- Massey, Z. B., Wong, N. C. H., & Barbati, J. L. (2021). Meeting the (trans)parent: Test of parasocial contact with transgender characters on reducing stigma toward transgender people. *Communication Studies*, *72*(2), 232–250. <https://doi.org/10.1080/10510974.2021.1876125>
- McDonald, M., Porat, R., Yarkoney, A., Reifen Tagar, M., 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>
- Milkman, K. L., Gromet, D., Ho, H., Kay, J. S., Lee, T. W., Pandiloski, P., Park, Y., Rai, A., Bazerman, M., Beshears, J., Bonacorsi, L., Camerer, C., Chang, E., Chapman, G., Cialdini, R., Dai, H., Eskreis-Winkler, L., Fishbach, A., Gross, J. J., ... Duckworth, A. L. (2021). Megastudies improve the impact of applied behavioural science. *Nature*, *600*, 478–483. <https://doi.org/10.1038/s41586-021-04128-4>
- Montiel, C. J., de la Paz, E., & Cerafica, Z. I. (2019). (De) humanization and trust in an asymmetric Muslim–Christian conflict: Heroes, Kafirs, and Satanas. *Peace and Conflict: Journal of Peace Psychology*, *25*(4), 300–311. <https://doi.org/10.1037/pac0000418>
- 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/10.1177/194855062111012793>
- Moore-Berg, S. L., Hameiri, B., Falk, E., & Bruneau, E. (2023). Reducing Islamophobia: An assessment of psychological mechanisms that underlie anti-Islamophobia media interventions. *Group Processes & Intergroup Relations*, *26*, 555–578. <https://doi.org/10.1177/13684302221085832>
- Moyer-Gusé, E., Mahood, C., & Brookes, S. (2011). Entertainment-education in the context of humor: Effects on safer sex intentions and risk perceptions. *Health Communication*, *26*(8), 765–774.
- Nabi, R. L., Moyer-Gusé, E., & Byrne, S. (2007). All joking aside: A serious investigation into the persuasive effect of funny social issue messages. *Communication Monographs*, *74*(1), 29–54. <https://doi.org/10.1080/03637750701196896>
- Nir, N., & Halperin, E. (2019). Effects of humor on intergroup communication in intractable conflicts: Using humor in an intergroup appeal facilitates stronger agreement between groups and a greater willingness to compromise. *Political Psychology*, *40*(3), 467–485. <https://doi.org/10.1111/pops.12535>
- Paolini, S., White, F. A., Tropp, L. R., Turner, R. N., Page-Gould, E., Barlow, F. K., & Gómez, Á. (2021). Intergroup contact research in the 21st century: Lessons learned and forward progress if we remain open. *Journal of Social Issues*, *77*(1), 11–37. <https://doi.org/10.1111/josi.12427>
- Pavetich, M., & Stathi, S. (2021). Investigating antecedents of Islamophobia: The role of perceived control over terrorism, threat, meta-dehumanization, and dehumanization. *Journal of Community & Applied Social Psychology*, *31*(4), 369–382. <https://doi.org/10.1002/casp.2512>
- Prati, F., Crisp, R. J., & Rubini, M. (2015). Counter-stereotypes reduce emotional intergroup bias by eliciting surprise in the face of unexpected category combinations. *Journal of Experimental Social Psychology*, *61*, 31–43. <https://doi.org/10.1016/j.jesp.2015.06.004>
- Saguy, T. (2018). Downside of intergroup harmony? When reconciliation might backfire and what to do. *Policy Insights From the Behavioral and Brain Sciences*, *5*(1), 75–81. <https://doi.org/10.1177/2372732217747085>

- Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs*, 72(1), 92–115. <https://doi.org/10.1080/0363775052000342544>
- Stathi, S., Husnu, S., & Pendleton, S. (2017). Intergroup contact and contact norms as predictors of postconflict forgiveness. *Group Dynamics: Theory, Research, and Practice*, 21(1), 20–39. <https://doi.org/10.1037/gdn000060>
- Steizinger, J. (2018). The significance of dehumanization: Nazi ideology and its psychological consequences. *Politics, Religion & Ideology*, 19(2), 139–157. <https://doi.org/10.1080/21567689.2018.1425144>
- Tapley, R. (2006). The value of humor. *Journal of Value Inquiry*, 40(4), 421–431. <https://doi.org/10.1007/s10790-006-9007-y>
- Varma, A. (2020). Evoking empathy or enacting solidarity with marginalized communities? A case study of journalistic humanizing techniques in the San Francisco Homeless Project. *Journalism Studies*, 21(12), 1705–1723. <https://doi.org/10.1080/1461670X.2020.1789495>
- Vezzali, L., Visintin, E. P., Bisagno, E., Bröker, L., Cadamuro, A., Crapolicchio, E., De Amicis, L., Di Bernardo, G. A., Huang, F., Lou, X., Stathi, S., Valor-Segura, I., & Harwood, J. (2022). Using sport media exposure to promote gender equality: Counter-stereotypical gender perceptions and the 2019 FIFA Women’s World Cup. *Group Processes & Intergroup Relations*, 26, 265–283. <https://doi.org/10.1177/13684302221075691>
- Zhang, Y., & Zinkhan, G. M. (2006). Responses to humorous ads: Does audience involvement matter? *Journal of Advertising*, 35(4), 113–127. <https://doi.org/10.2753/JOA0091-3367350408>

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher’s web site:

Figure S1. Study 2 unstandardized path model assessing the effect of the Mean Tweets condition on anti-Muslim policy support through dehumanization. * $p < .05$; ** $p < .01$; *** $p < .001$.

Figure S2. Study 2 unstandardized path model assessing the effect of the ‘Serious Mean Tweets’ condition on anti-Muslim policy support through dehumanization. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table S1. Study 1 Demographics of Participants Per Condition

Table S2. Study 1 Follow Up Assessment: Demographics of Participants Per Condition

Table S3. Study 2 Demographics of Participants Per Condition